Palacký University Olomouc
Faculty of Arts
Department of Psychology

Libet's experiment, its replicability, validity and clinical potential

Libetův experiment, jeho replikabilita, validita a klinický potenciál

Dissertation thesis

Author: Mgr. Tomáš Dominik

Supervisor: PhDr. Mgr. Roman Procházka, Ph.D.

Olomouc

2019

Words of thanks

Within the circles of experimental and cognitive psychologists, Libet's experiment is sometimes dubbed "the experiment which started a thousand careers". I sincerely hope that this thesis will add one more to that thousand. Present dissertation thesis reflects more than three years of research I, together with my colleagues, have conducted in this field. Since the beginning of my undergraduate studies, I was intrigued by the question of free will, but it took me four years to take the courage to tackle Libet's experiment because I viewed it as something far beyond my capabilities. Nonetheless, thanks to encouragement of my supervisors and thanks to my luck of meeting the right researchers sharing my interests, I can now address some questions persisting in the "libetian" discussions to this day. I would like to express my gratitude to all the people who allowed me to conduct this research, namely:

- Roman Procházka for his willingness to supervise this thesis, to share his clinical expertise with me and to read the manuscript of this thesis.

- Martin Zielina for his supervision of my previous master thesis, for the stimulating discussions we had throughout my Ph.D. studies and for reading the manuscript of this thesis.

- Daniel Dostál for mental stimulation, his inspiring zeal for cognitive psychology and data science, for help with data analyses and for reading the manuscript of this thesis.

- Daniel Dostál, Martin Zielina, Roman Procházka, Jan Šmahaj and Zuzana Vaculčíková Sedláčková for cooperating on our projects and papers.

- Barbora Sedláková, Markéta Zajícová and others who allowed me to collect data within their institutions.

- Michal Kluka for language help with Studies 1 and 2.

- Miroslav Kuba and Jan Kremláček for consultations pertaining the ERP analyses.

- Patrick Haggard, Axel Cleeremans and all the members of their labs for hosting me and encouraging me to join the flow of current neuroscience and cognitive sciences.

- The Department of Psychology, Faculty of Arts at Palacký University Olomouc, and especially its head Matúš Šucha, for creating an ideal academical working environment.

- My family and especially my girlfriend, Jana Mišelnická, for much needed support.

I hereby proclaim that I have written this dissertation thesis named "Libet's experiment, its replicability, validity and clinical potential" by myself, under the supervision of PhDr. Mgr. Roman Procházka, Ph.D. and using only cited literature. Studies 1 and 2 presented in this thesis were published and are used with consent of all co-authors (see Appendix 3).

Místopřísežně prohlašuji, že jsem dizertační práci na téma: „Libetův experiment, jeho replikabilita, validita a klinický potenciál" vypracoval samostatně pod odborným dohledem PhDr. Mgr. Romana Procházky, Ph.D., a uvedl jsem všechny použité podklady a literaturu. Studie 1 a 2 prezentované v této práci byly již publikovány a jsou použity se souhlasem všech spoluautorů (viz Appendix 3).

In Olomouc on the 18th June 2019   Signature ....

Table of contents

Introduction

Libet's experiment is known to many as the experiment which denied free will. It would be tempting to introduce this thesis by a discussion of whether we do or do not have free will and what either of these options mean for our everyday lives. However, the present thesis does not study Libet's experiment from this perspective. When Libet et al. published their results in 1982 and 1983, free will was not even mentioned in their papers, suggesting that it is important to investigate the findings independently of what their implications might be. In my research, I have followed Libet's experiment itself, not its implications, because implications of any research are based on whether this research is considered reliable and valid, which remains in question in the case of Libet's experiment.

There are three studies presented in the thesis, connected to its three objectives. The first objective is to scrutinize one of many debated features of Libet's experiment and test its validity in a novel way. The second objective is to analyse the experiment as a whole and—by replicating it in a complex way—investigate the robustness of its results. The third objective is to explore its clinical potential, i.e. assess whether the experiment can be utilized in the setting of clinical psychology. The reader might wonder why I would intent to discuss a clinical use of an experiment, validity of which was not agreed upon for more than 30 years. The answer is that several studies exploring the implications of Libet's experiment produced interesting results suggesting that certain parts of the original methodology might bear this potential. I believe that it is worth investigating such possibility.

This thesis is divided into theoretical and empirical part. **The theoretical part** deals with Libet's experiment in general. In this section, I (1) describe the historical background of Libet's experiment, (2) introduce Libet's methodology, results and interpretations, and (3) discuss the experiment from other researchers' point of view. The purpose of the theoretical part is to introduce the original experiment, since it is the cornerstone of the present thesis.

The empirical part consists of three empirical studies. Study 1 presents an investigation of validity of obtaining introspective reports of measuring the intention to move using a device sometimes called Libet's clock, which is one of the key points in Libet's methodology. This study was published in the first year of my Ph.D. study (see Dominik et

al., 2017). Study 2 presents methods, results and discussions of our replication study, since a replication of Libet's experiment in all its complexity has not been conducted before that. This study was published in 2018 (Dominik et al., 2018a, 2018b). Study 3 reflects the question of whether it is possible to use Libet's experiment, namely its introspective component, in clinical context as a psychodiagnostical method. In this study I explore its ability to discriminate between healthy and pathological populations, as well as its reliability.

Within the introduction, I would like to address a potential issue regarding similarities between this text and my previous master thesis (Dominik, 2016), which it **expands**. My master thesis contains descriptions of the replication methodology (Study 2), but—as we were in the middle of the data collection process at that time—it does not contain any information on data analysis, results, interpretations and discussions presented here and in Dominik et al. (2018a, 2018b). In fact, this part of the study constituted a significant part of the first and second year of my doctoral studies. The master thesis also contains a report of an experiment similar to the study of validity of introspective impressions presented in Dominik et al. (2017) and in Study 1 of this text. However, we conducted this experiment again after my master thesis was defended and used more sophisticated methods of data analysis. I do use some portions of my master thesis here, especially in Section 1, but do so in accordance with recommended guidelines and accepted practice at the Department of Psychology at Palacký University Olomouc.

Theoretical part: History of Libet's experiment, its methods and discussions

1 Historical background

The history of mankind is defined by exciting discoveries about the world around us—we invented the wheel, fire, agriculture, metalworking; we have circumnavigated the Earth, we have learnt to fly planes, we have used spaceships to visit the Moon. We know a lot about the world around us; too much in comparison with what we know about our own brains. It was at the turn of the 19th and 20th century, when our thoughts about mental processes surpassed the borders of philosophy and entered the realm of empirical science.

Brain research set off in several directions. One of these directions was research utilizing brain stimulation. In 1937, for instance, Penfield and Boldrey (in Schott, 1993) published a graphical "brain map" illustrating how peripheral parts of human body are represented and organized in the brain cortex; this representation is nowadays known to every psychology student as the "homunculus". Penfield and Boldrey found that by electrical stimulation of specific brain regions, they can elicit various kinds of perceptual experiences and motoric activity. However, this way of studying human brain is inefficient and sometimes questionable because it requires the researchers to penetrate subject's skull or remove a part of it and physically intervene with the brain tissue below; not speaking about the fact that participants of such studies are almost exclusively neurological patients with disorders severe enough to require such an invasive surgery. Fortunately, at the end of the 19th century, a different approach called electroencephalography emerged, becoming precursor of modern neuroimaging methods.

The electroencephalogram (EEG) was first described in animals by Richard Caton in 1875 (Andreassi, 2007; Haas, 2003). Using this method in humans, Hans Berger revolutionized brain research by discovering that synchronized electrical activity of neurons in the brain can be analysed in terms of frequency and proposed that there are at least two distinct brain states—one exhibiting waves with lower frequency (called **alpha waves**) and the other exhibiting waves with higher frequency (called **beta waves**) (Andreassi, 2007; Haas, 2003). Later, additional frequency bands were identified and ever improving

techniques of EEG frequency analysis are still proving useful today. However, in the second half of the 20th century, an alternative approach to EEG data analysis emerged, which is now known as the event-related potentials (ERPs, also known as evoked potentials, see Andreassi, 2007; Bareš, 2011). Unlike frequency analysis, which is useful in describing brain **states**, event-related potentials are informative in those kinds of situations, in which we are interested in brain activity responding to specific known **events** (be it a flash of light, an audible beep or a toe movement). Since the middle of the 20th century, many types of ERPs were identified, defined and named—for instance, a P300 wave is known to be elicited after an onset of an attended stimulus (see Picton, 1992) and a contingent negative variation (CNV) was shown to emerge between "get ready" and "go now" stimuli (Walter, Cooper, Aldridge, McCallum, & Winter, 1964; Sanquist, Beatty, & Lindsley, 1981).

The main theme of the present thesis is Libet's experiment and the most relevant ERP related to it is the **readiness potential** (RP), first discovered and described by Kornhuber and Deecke (1965). According to these authors, the readiness potential (called *Bereitschaftspotential* in their original paper) is a gradual negative shift in the EEG voltage preceding a self-induced movement. Authors reported it to increase in amplitude with growing intentional engagement of the individual. Additionally, they also noted that no temporal relation between the readiness potential and alpha waves phase was found, hence dismissing potential objection that the RP waveform can be in fact just a specific manifestation of the alpha rhythm.

One could expect that this is where Benjamin Libet entered the scene of neuroscientific brain research. However, even though Libet is most known for utilization of the RP recordings, his original methods were in fact **invasive brain-stimulation experiments**. As Libet states in his monograph (Libet, 2004), his scientific career begun with studying absolute thresholds of conscious experiences in cooperation with Dr. Bertram Feinstein. He addressed this topic throughout 1960s and 1970s. Libet's most significant finding from that period is that to elicit a conscious somatosensory experience, a continuous train of cortical stimulation in duration of at least 500 ms is needed. This is in striking contrast with the fact that stimulation at the periphery (e.g. on the skin) requires durations, which are negligible in comparison to said 500 ms (see Libet, Alberts, Wright, & Feinstein, 1967). This finding inevitably leads to the following question: how is it possible that individuals perceive touch on the skin and own realization of the touch as simultaneous, if the brain requires 500 ms of

constant stimulation to elicit such realization? As Libet (2004) explains, it is possible that all conscious experiences are referred to retrospectively (this concept is known as the **backward referral** hypothesis). Libet supported the interpretation by findings of other authors; for instance, Crawford (1947 in Libet, 2004) found that a tiny flash of light followed within 100 ms by a large light flash is not consciously perceived, even though it is in itself supraliminal. Libet even extended this phenomenon by demonstrating that the masking stimulus can be delivered directly to the brain cortex or that the second stimulus might not only mask, but even magnify intensity of the first stimulus (Libet, 2004). It should be noted that backward referral interpretation of Libet's early findings is not without controversy (see Gomes, 1998; Gomes, 2002).

After death of Dr. Feinstein, Libet with his colleagues turned their attention to studying volitional actions (Libet, 2004). This research interest ultimately led to what is now being referred to as "the famous experiment on free will" in the popular culture. I should clarify here that when I refer to "Libet's experiment" in this text, I always have this particular experiment in mind. Libet followed up on Kornhuber and Deecke's discovery of the readiness potential and asked the question how conscious experience of the movement relates to pre-motor activity in the brain. More specifically, Libet and his team extended the assumption that there is approximately 500 ms latency in the process of generating conscious experience (see Libet et al., 1967). One could therefore assume (which Libet probably did, see Libet, 1985, p. 536) that similar latency may apply even for conscious experiences connected to planning and executing a motor action. In the next chapters, we will explore this experiment in greater detail.

2 Methods and results of Libet's experiment

The main objectives of Libet's experiment can be summarized as follows: (1) to explore the characteristics of the RPs preceding a voluntary movement (Libet, Wright, & Gleason, 1982) and (2) to compare the timing of the RP onsets to the subjective experience of preparing to move (Libet, Gleason, Wright, & Pearl, 1983). The most well-known outcome of the experiment is that the onsets of RPs precede subjective realization of an intention to act. These statements summarize what Libet's experiment is generally known for but miss several aspects important for its implications. All topics covered in this thesis relate to Libet's experiment in some way. Therefore, it is important to introduce its methodology and

results in detail, so that the reader is familiar with all important findings and features of the experiment.

Before I proceed to rather technical descriptions of Libet's methodology, however, I should answer one important question: **why does Libet's experiment even matter so much**? The obvious answer might be that it brought important findings about human nature and the nature of free will. Nevertheless, this answer is somewhat hardly defendable, given that virtually every single aspect of Libet's experiment has been deeply scrutinized, which lead to plenty of doubt about its conclusions with respect to the problem of free will. I would object that Libet's experiment was never primarily about free will. The first three papers published on the subject (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, Wright, & Gleason, 1983) did not even contain the phrase "free will". To the question above I would reply that Libet's experiment might not give us conclusive evidence for how volition works, but it showed us in truly pioneering fashion the way to study volition empirically. Besides the impact on the research in volitional processes, however, Libet's ideas had also consequences in other areas of psychological research. In 1977, a highly influential paper was published showing how unreliable and inaccurate introspective reports about causes of one's own actions are (Nisbett & Wilson, 1977). Libet was important in this context because even in the atmosphere of high scepticism towards introspective methods, he managed to find a way to take introspective reports of his participants into serious account, which was endorsed by contemporary commentators, such as Doty (1985). While being aware that these kinds of introspective reports were not exactly what Nisbett and Wilson had in mind, I would like to express my opinion here that the introspective evidence (albeit easily biased) should never be completely excluded from psychological science.

In the following pages, I will introduce the reader to Libet's methodology in all its complexity. To do this, I researched four primary papers (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, Wright, et al., 1983; Libet, 1985). In this regard, I should dedicate a few lines to suggest possible reasons why Libet published the experiment in four separate articles. The answer may be that the authors simply intended to separate different types of conclusions, as the experiment is notably complex. The first study, published by Libet et al. (1982), emphasized the analysis of the readiness potentials. They showed, among other findings, what the RP looks like if it occurs before a spontaneous movement compared to a pre-planned movement and that it does not occur before a skin stimulus. The second paper,

published by Libet, Gleason, et al. (1983), introduced the introspective reports such as the moment of the first conscious urge to move (called "W" as in "wanting") or the subjective impression of the actual initiation of the movement (called "M" as in "movement"). The authors pointed out that the RP onsets generally precede not only the movement itself, but also the conscious awareness of wanting to move. The third paper (Libet, Wright, et al., 1983) reports a smaller study supplementary to the original one, but I consider it to be part of the whole narrative, as it was conducted on mostly the same participants in almost identical setting; its importance lies in the introduction of the concept of a conscious veto, which is an important feature of the experiment's interpretation. I should point out that we have, in mistake, claimed in our paper (Dominik et al., 2018a) that the first mention of the conscious veto was presented in Libet (1985); I would like to set the record straight here and state that it was introduced two years earlier in Libet, Wright, et al. (1983). However, to be completely precise, a suggestion of a conscious veto (although not yet fully developed) can be found also in Libet, Gleason, et al. (1983). The fourth paper (Libet, 1985) contains mainly discussions and interpretations of the findings, as well as some comments by other researchers at that time.

In this chapter, I will limit my comments on the original experiment to a necessary minimum, as I would like to first introduce it from Libet's point of view. Critique and discussion of the methodology can be found in chapters 2.5, 3, 4 and 5, which point to several follow-up studies conducted by other researchers and ourselves.

2.1 Libet's sample

Libet's participants in the main study were 6 university students (5 female, 1 male; all right-handed). Data of one female participant could not be fully analysed due to poor EEG quality and minimal RP amplitude. This sample was divided into two subsamples. Subsample 1 consisted of three females, subsample 2 consisted of two females and one male. Sessions with participants in the subsample 2 were conducted several months after the sessions with subsample 1; some methodological improvements and changes based on experiences with subsample 1 were employed in session with subsample 2. Some data collected in the pilot trials with Libet himself were identified as suitable for presentation and were included in the analyses as well (Libet et al., 1982, p. 323; Libet, Gleason, et al., 1983, p. 624)

2.2 Libet's equipment

Libet's design requires electrophysiological laboratory. In the most general sense, four separate systems are needed to conduct the experiment: an **electroencephalograph** (recording electrical activity of the brain), an **electromyograph** (recording electrical activity of muscles), a **skin stimulator** and a computer with specific **clock software** for collecting introspective data (Libet et al., 1982; Libet, Gleason, et al., 1983). Initially, Libet also employed electrooculographic recordings to reduce the number of artefacts caused by eye movements and blinking but discontinued the measurements, because such artefacts were only rarely found (Libet et al., 1982, p. 323).

2.2.1 Electroencephalographic recordings (EEG)

The purpose of EEG in Libet's experiment is to record cognitive event-related potentials. For the recordings, Libet utilized a DC system using amplifiers with low-pass filter set to 35 Hz. This means that virtually all signal components with frequency higher than 35 Hz were filtered out from the recording (for more information about filtering, see Stern, Ray, & Quigley, 2001, pp. 41–43). Libet used electrodes made of alloy of silver and silver chloride (Ag/AgCl), which is a commonly used material for physiological electrodes (see Stern et al., 2001, p. 37). Electrodes were attached to skin cleaned with acetone using an adhesive electrode paste (Libet et al., 1982, p. 323).

Most electrodes were applied according to standard 10-20 system (see Figure 1) on these locations: vertex (C_z), left parietal area (P_3), left prefrontal area (F_{p1}) and right prefrontal area (F_{p2}). Besides these four locations, Libet also placed two electrodes to non-standard sites: the contralateral (C_c) and ipsilateral (C_i) electrodes were located on the prerolandic (i.e. precentral) motor areas contralaterally and ipsilaterally to the right hand, which was used to execute the studied movement. A ground electrode was positioned on the left earlobe (Libet et al., 1982, p. 323) and reference electrodes were placed on mastoids (Libet, Gleason et al., 1983, p. 624). Libet also used additional large ground electrode on the arm when using the skin stimulator (Libet et al., 1982, p. 323).
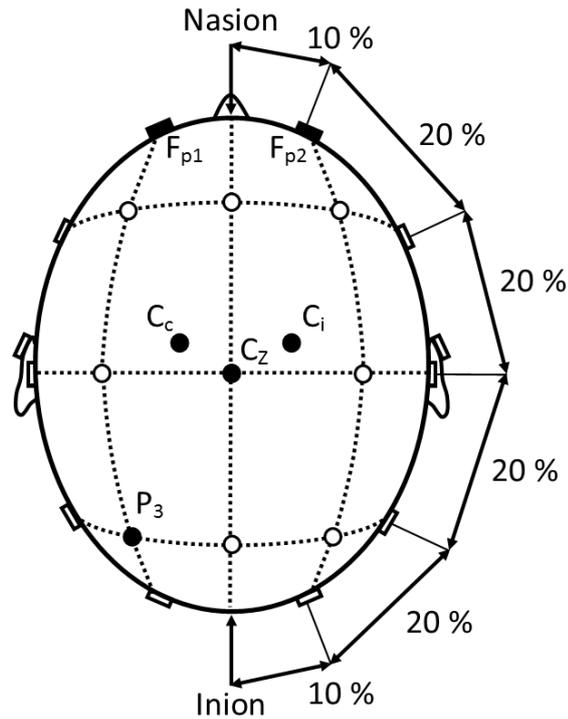
Figure 1: EEG electrodes sites on participants scalp in Libet's original experiment. Electrodes F_{p1}, F_{p2}, C_z a P_3 are placed within standardized 10-20 system, whereas electrode C_c and C_i were added in accordance with the experiment design.

To analyse the signal, Libet and his colleagues extracted 2000 ms long epochs starting 1400 ms before an event of interest (i.e. a movement or stimulus delivery) in each trial. There were always 40 consecutive trials in a series, and the epochs from these 40 trials were subsequently averaged to reduce the noise otherwise obscuring the RP waveform. The voltage baseline was adjusted before each trial (Libet et al., 1982, p. 324; Libet, Gleason, et al., 1983, pp. 624–625).

2.2.2 Electromyographic recordings (EMG)

To register muscle movements, a bipolar EMG from the right forearm was recorded with amplifier set to 3 kHz low-pass filter. When EMG amplitude reached an activation value, a computer recorded the time of EMG activation and used it as a synchronizing event for the clock software. The participants were instructed to perform a flexion of a wrist or fingers, while the movement was required to be fast enough, so that EMG reaches the activation value in 10–20 ms. Participants were trained to do this using auditive biofeedback, transforming EMG signal into sound played in a loudspeaker (Libet et al., 1982, p. 323; Libet, Gleason, et al., 1983, p. 624).

2.2.3 Experimental clock

Libet's experimental clock (sometimes also referred to as the "rotating spot method", see Pockett & Miller, 2007; Miller, Vieweg, Kruize, & McLea, 2010) has an interesting history, as it was invented long before Libet utilized it in his research. Its development started with Wilhelm Wundt, who was interested in interindividual differences in the ability to perceive time based on an object moving though individual's field of view (Cairney, 1975). According to Cairney, the clock was originally used in methodologies called "complication experiments". In these, two competing stimuli of different modalities were presented, while the participant was instructed to attend one of them and subsequently state when he registered either the attended or the unattended (so called "complication") stimulus.

Figure 2a: The original Libet's clock   Figure 2b: Libet's clock used in our studies
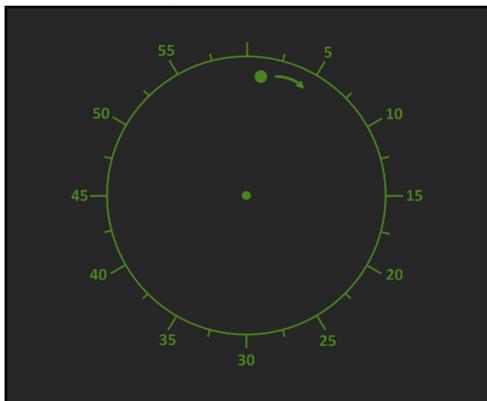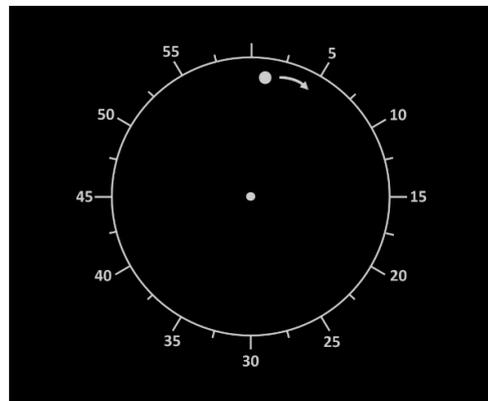
Figure 2: Schematic depiction of Libet's experimental clock. The arrow signifies the direction of the spot movement and was not actually displayed. Figure 2a resembles our approximation of its form according to Libet's description. Figure 2b with increased contrast and white outline depicts the clock design used in Studies 1, 2 and 3 of this work.

Libet's version of the clock (which I will call "Libet's clock" in this text) was designed to allow his participants to report timing of subjective experiences retrospectively. The idea is that the participant reports the clock position in the moment of a certain subjective event. The clock face is in principle like regular clock, but the clock hands are replaced by a single, rapidly revolving dot. To construct the clock in 1980s, Libet and his team used a *cathode-ray oscilloscope* (CRO) with a circular display with 5 inches in diameter. During the experiment, the oscilloscope was placed 1.95 m in front of the participant, so that the viewing angle corresponded to approximately 1.8°. A dot revolved around the clock face, completing one whole revolution in 2.56 s. The circumference of the clock face was outlined by a circle and marked by numbers and radial lines perpendicular to the circle (see Figure 2a). The distance between two neighbouring radial lines (one longer and the other

shorter) would on regular clock correspond to 2.5 s. On Libet's version of the clock, it corresponded to 107 ms of real time (see Libet, Gleason, et al., 1983, p. 625). A fixation point was displayed in the middle of the clock.

2.2.4 Skin stimulation

In certain trials, a skin stimulus had to be delivered, mostly with the purpose to estimate participants' accuracy when reporting experiences using the experimental clock. The skin stimulus was electrical and was delivered via a pair of electrodes placed on the back of participant's hand. The voltage intensity was set as 115–120% of a threshold value (i.e. minimal registered intensity) of a stimulus lasting for 0.5 s. According to the authors, the stimulus intensity *"was thus sufficiently weak to make recognition somewhat difficult, but sufficiently above threshold to eliminate equivocation about stimulus delivery"* (Libet et al., 1982, p. 324).

2.3 Tasks in Libet's experiment

To fully understand the original experiment's methods, terms *session*, *series* and *trial* need to be clarified. Each participant attended 6–8 half-day *sessions*. Intervals between two following sessions were about one week long. The first and, in some cases, the second session were dedicated to training (*training sessions*); others were called *regular sessions*. Each of Libet's sessions consisted of a certain number of *series* of various types, which consisted of 40 individual *trials* each. Each series was also preceded by 10 retraining trials (Libet et al., 1982, p. 325). A trial was the lowest-level unit of the whole experiment and included an event (i.e. a skin stimulus or EMG activation) and a report of timing of certain subjective experience (reporting was excluded from the pre-set series, see below). EEG was continuously recorded throughout the session.

During the session, the participant was seated in a lounge chair in slightly reclining position. Each trial was initiated by a brief auditory signal, after which the participant was supposed to relax muscles of the head, neck and forearm, and blink a few times, if needed. After that, the participant was asked to fixate his or her gaze on the point in the middle of the oscilloscope. Libet provided a specific instruction regarding blinking—the participants were asked not to blink during the trial, but if an urge to blink arose, the participant could do so, but was asked to wait at least one whole revolution of the clock before making any further actions (Libet et al., 1982, p. 324; Libet, Gleason, et al., 1983, p. 625).

I will now address types of tasks present in Libet's methodology. Libet classified the tasks (and series, respectively) into categories but used different—and often confusingly inconsistent—labels throughout the primary sources. I would like to unify the terminology and create a system of labels, which will be used further in the text. I will present the labels first and then explain in detail the tasks they denote.

In Libet et al. (1982), the series were divided into three types: (1) *Self-initiated voluntary acts*, (2) *Pre-set motor acts* and (3) *Skin stimuli at unknown times*. The S*elf-initiated voluntary acts* series correspond to the M and W series introduced in Libet, Gleason, et al. (1983). However, because the M and W tasks are not completely identical, I will use the **M** and **W** labels to distinguish between the two procedures. *Pre-set motor acts* are introduced in Libet et al. (1982), further discussed in Libet, Wright, et al. (1983) and Libet (1985), but not mentioned in Libet, Gleason, et al. (1983). Even more confusingly, in Libet, Wright, et al. (1983), the series with the pre-set motor acts are called "M series", which is inconsistent with the use of the term "M series" in Libet, Gleason, et al. (1983). To deal with this inconsistency, I will label the pre-set series without vetoing the movement as **P** series. An additional subtype of pre-set series "with conscious veto" is discussed in Libet, Wright et al. (1983) and Libet (1985, p. 538) and I will label this variant as **Pv**. *Skin stimuli at unknown times* series—also called S series in Libet, Gleason, et al. (1983)—will be labelled **S** in this text. Even the S series are inconsistent throughout the papers, however, as Libet, Wright, et al. (1983) use the label "S series" for a combination of skin stimulation with pre-set methodology. I will label this variant of the S series as **Sp**.

2.3.1 Self-initiated voluntary acts (M and W series)

In both the M and W series, the participants were asked to conduct a flexion of a wrist or fingers whenever they felt like doing so (see Libet et al., 1982, p. 324; Libet, Gleason, et al., 1983, p. 625). The participants received an instruction emphasizing that the movement should be spontaneous and not pre-planned (this was not included in the instruction in the first half of the sessions in subsample 1; see chapter 2.1 for clarification of the subsamples). In most of the series, the experimenters retrospectively asked about whether any pre-planning occurred or not.

The difference between the M and W series was in the subjective experience reported after conducting the movement. In the **M task**, the participant is asked a few seconds after the movement to report the earliest time of when he or she felt that his or her hand moved.

In the **W task**, the participant is asked to report the first *wanting* to move (Libet also stated alternative terms for this experience such as *urge*, *intention* and *decision*, see Libet, Gleason, et al., 1983, p. 627). Within the W series, the experimenters also asked about a feeling of being surprised by the movement.

In both the M and W tasks, the experimenters emphasized that the reporting phase will come after the movement and that the participants should not worry about it in advance. Despite this, a concern might arise that this request to report the time of a subjective experience of any kind might distort recorded EEG potentials. Nevertheless, Libet argued that the readiness potential recorded in these series did not differ from the potentials recorded in series in which no reports were made (Libet et al., 1982, p. 325).

Every M and W series contained 40 trials and were conducted mainly within the first four regular sessions; in later sessions, the M series were conducted only occasionally. Between each two consecutive sessions with the same participant, the order of M and W series was alternated.

2.3.2 Pre-set motor acts (P and Pv series)

The P and Pv series were not included in the first four regular sessions in either of the subsamples and were added to the experiment only in later sessions and in the supplementary study reported in Libet, Wright, et al. (1983). In the P series (see Libet et al., 1982, p. 325), the participants were asked to conduct a rapid flexion of a wrist or fingers (just as in the M and W series) in certain moment, which was set beforehand on the experimental clock. Specifically, the participant was told a certain "pre-set" clock position and was supposed to execute the movement when the clock reached the position. These series contained four blocks of ten trials. In each block, the pre-set position was different (10, 20, 40 and 50 "seconds"). Before any movement, the clock had to complete at least one revolution. The participants were asked to conduct the movement as close to the pre-set time as possible. Their accuracy was measured by EMG recordings compared to the actual pre-set time. A feedback on the accuracy was given to the participant at the end of every series of 40 trials.

The Pv series used a similar setup. The participant was given a prearranged time with the instruction to prepare to conduct the movement just as in the P series, but then "veto" the intention approximately 100–200 ms before the pre-set time (Libet, Wright, et al.,1983, p. 369; Libet, 1985, p. 538). Libet is not completely clear on further details on the Pv series,

but we can assume that the rest of the methodology corresponded with the P series (e.g. four blocks by ten trials each).

2.3.3 Skin stimuli at unknown times (S series)

Unlike in the M, W, P and Pv series, a skin stimulation was employed in the S series (see Libet et al., 1982, p. 325; Libet, Gleason, et al., 1983, p. 625). The stimulus was delivered via an electrode attached to the back of participant's right hand. The experimenter sent slightly supraliminal electrical pulse to the hand during the second or third clock revolution, while the participants were instructed that the stimulus can be delivered at any time after the first revolution. The participants' task was to report when they registered the stimulus, just as they reported the experiences in the M and W series. After each series of 40 trials, the participants received a feedback on the accuracy of their reports. If the reports deviated significantly from the actual time of the stimulus delivery, this deviation was considered participant's specific "bias" or "shift" (Libet, Gleason, et al., 1983, p. 627–628).

In addition to normal "tested" S series, each session was preceded by a short S series of 25 trials intended to re-familiarize the participant with the introspective principles (Libet, Gleason, et al., 1983. p. 628). The exceptional attribute of these training S series was that the participants were given a feedback every 5 trials on how accurately they reported the S time.

2.3.4 Skin stimuli at pre-set times (Sp series)

The Sp series were introduced in Libet, Wright, et al. (1983). Just as in the S series, skin stimuli were used here in the same technical setup (reasonably supraliminal intensity, electrode on the back of the right hand etc.), but with two important differences. First, the stimuli were delivered at pre-set times known to the participant in advance (with the 10, 20, 40 and 50 "seconds" timing, just as in the P and Pv series), and so the participant was not required to report the time of the stimulus delivery afterwards. Second, to keep the participant attentive to the stimuli, in 5–8 of the 40 trials in total, the stimulus was omitted and the participant was asked to count the number of the omitted stimuli in total (Libet, Wright, et al., 1983, p. 368).

In the M, W and S series, a time of a subjective experience was reported 500–800 ms after EMG activation or stimulus delivery (this interval is referred to as the *continuation*

interval because the dot on the clock's circumference continued moving during this period of time). The reports could be provided in two distinct ways. Libet called these ways of reporting *modes of recall* and distinguished an ***absolute*** **(A)** mode and an ***order*** **(O)** mode (see Libet, Gleason, et al., 1983, p. 626). The A mode is a simple and straightforward approach—the participant is asked to state the clock position present when he or she registered the subjective event (either a movement in the M series, intention to move in the W series, or stimulus delivery in the S series).

The O mode of recall was more sophisticated and was perceived by the participants to be somewhat easier to use than the A mode. After the continuation interval ended, the moving dot "jumped" to specific *stop time* position. The stop time position was in each of the 40 trials randomly chosen from 41 evenly distributed positions in an interval called *stopping range*, which spanned from 400 ms before the event to 200 ms after it (1 of the 41 possible stop time positions was omitted in each series). The participant was asked to report the time of an event relative to current stop time. In this regard, three possibilities could occur: (1) the reported event occurred before the stop time, (2) after the stop time or (3) precisely at the stop time. The mean time of a reported experience was calculated based on the following formula (Libet, Gleason, et al., 1983, p. 629):

(upper, positive end of "stopping range") – (time interval between "stop times") × (number of points – ½)

Number of points was determined based on participant's reports. For each report of the event occurring before the stop time, the participant was given 1 point, for each report of the event occurring precisely at the stop time, the participant was given ½ of a point, and for each report of the event occurring after the stop time, the participant received no points.

2.4 Libet's results

There are two categories of results in Libet's experiment, which might justify its publication in the two main papers (Libet et al., 1982; Libet, Gleason, et al., 1983). One category consists of descriptions and elaboration of the readiness potentials (see Libet et al., 1982). The other category are results pertaining subjective timing of mental events—specifically, the reports of the movement initiation (M reports), intention to move (W reports) and stimulus delivery (S reports) (Libet, Gleason, et al., 1983). I will divide this chapter according to this distinction.

2.4.1 Results of neural data analyses

Before I begin presenting Libet's results regarding the readiness potentials and their interpretations, I should explain the general idea behind RP analysis. I have already stated in chapter 2.2.1 that EEG data were averaged (see also Libet et al., 1982, p. 324). The reason for this is that event-related potentials in general tend to have rather small amplitude compared to ever-present EEG background noise (Andreassi, 2007, pp. 127–128). Segmenting the EEG data into time intervals called *epochs*—which are time-locked to a certain event (such as stimulus delivery or movement initiation)—allows us to average multiple epochs so that most of the random noise get averaged-out. This way, only the event-related potentials, which presumably occur in all epochs and have more or less stable characteristics, remain in the data together with only a small portion of noise.

Libet and his colleagues (1982) arrived at the conclusion that the RPs they found could be classified into three types called simply RP I, RP II and RP III. **The type I RP** should have a distinct onset more than 700 ms before the EMG activation, while average reported onset was in time -1055 ms (meaning 1055 ms before the relevant event, which is the EMG activation in this case). The RP I waveform was reported have a "ramp-like" shape (Libet et al., 1982, p. 326; see Figure 3a). Authors reported that the onset was exceptionally early (-1400 ms) in those trials in which spontaneity was not mentioned in the instructions, presumably allowing the participants to pre-plan their movement more often. Authors asked the participants after most of the series whether they were pre-planning their actions or not and reported—while noting that this finding is inconclusive—that there might be some connection between reported pre-planning of the movement and RP I occurrence. Conversely, when the participants reported that the action was spontaneous, no RP I was detected, with one borderline type I / type II exception (Libet et al., 1982, pp. 327–329).

Type II RP was reported to have its onset somewhere between -400 and -700 ms (-577 ms on average in the subsample 2). RP II's waveform is *"somewhat dome-shaped"* (Libet et al., 1982, p. 326; see Figure 3b). Libet reported that even in the case of RP II, negative deflections earlier than 700 ms could occur, but these negativities were rather irregular, were low in amplitude and did not rise as steadily as RP I.

Type III RP is difficult to describe, as Libet did not elaborate on this category. Libet et al. (1982, p. 326) suggest that RP III should have onset between -250 and -200 ms and that it has only a small duration and amplitude (see Figure 3c). According to Table 1 in Libet et

al. (1982) and Table 2 in Libet, Gleason, et al. (1983), there were only 5 instances of type III RP across all series. Libet and his colleagues asked the participants whether they felt surprised by the action, and so it might be reasonable to ask whether this feeling of being surprised was connected to RP III occurrence; nevertheless, the relation seemed to be non-existent (Libet et al., 1982, p. 330).

Figure 3a: Ideal type I RP

Figure 3b: Ideal type II RP

Figure 3c: Ideal type III RP
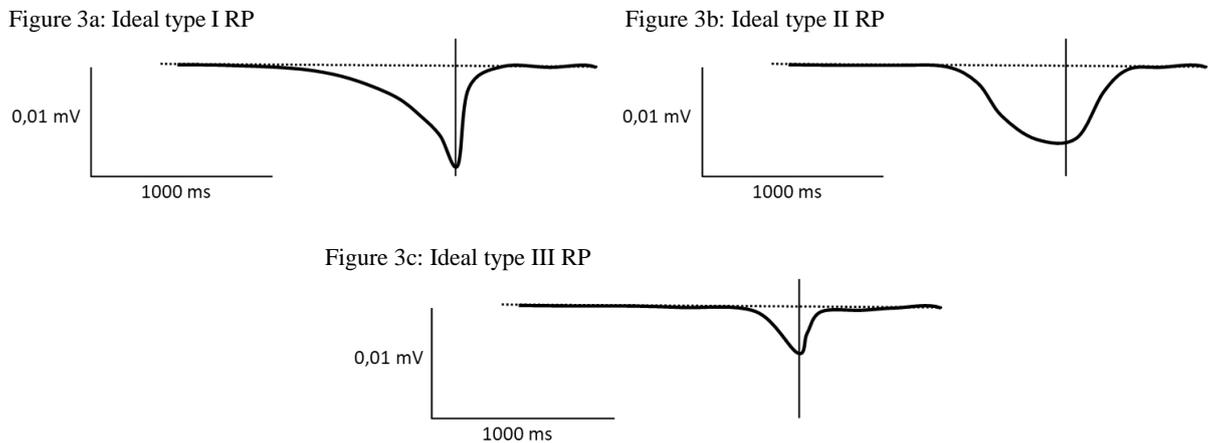
Figure 3: Depictions of ideal type I, type II and type III RPs. These figures do not represent any data, they are drawn by hand to illustrate the most characteristic features of the respective RP types according to Libet et al. (1982). The horizontal dotted line represents the baseline, the solid vertical line represents the EMG onset.

Interestingly, Libet and his colleagues also described a specific RP type beyond the types I, II and III terminology, and that is the readiness potential often obtained in the P series. They labelled this type of RP a **pre-set RP** and described it as similar to RP I, but with earlier onset and higher amplitude (Libet et al., 1982, p. 330). Simultaneously, Libet acknowledged similarities between pre-set RPs and CNV component found before the S_2 stimulus in the S_1-S_2 paradigm (S_1 being a warning stimulus and S_2 being a "go" stimulus; see Walter et al., 1964; Sanquist et al., 1981; Bareš, 2011). It should be noted that it was already suggested in the 1980s and 1990s that CNV, and especially its late component, might be related to the readiness potential (see Sanquist et al., 1981, p. 639; van Boxtel & Brunia, 1994). One could argue, however, that Libet's conjecture about connection between the pre-set RP and CNV cannot be justified, because the pre-set task is not a typical S_1-S_2 paradigm since there is no distinct warning stimulus. Nevertheless, Libet adds that shortening the time of the pre-set task from the clock starting to move to the moment it reached the pre-set position to about 1 second, the RP changed into a form closely resembling typical CNV waveform (Libet et al., 1982, p. 330). By this, Libet probably suggests that the start of the clock movement can be considered an S_1 event.

While discussing the pre-set RP, one could ask what the RP looks like in the Pv series, when the participant is instructed to prepare to make the movement, but then refrains from conducting it. Libet states that the vetoed movement was preceded by a ramp-like negativity, which differed (in most cases) from the regular pre-set RP by culminating somewhen between 150 and 250 ms before the clock reached the pre-set position and then remained flat or slowly begun returning to baseline (Libet, Wright, et al., 1983, p. 369; Libet, 1985, p. 536). Based on this finding, Libet suggests that it is possible for the human brain to build up a preparatory motor activity and subsequently interrupt such process.

Libet and his colleagues were especially interested in the timing of the various RPs. I have already stated the intervals of the onsets typical for RP I, II and III, but I did not explain how exactly these onsets were supposed to be assessed. There is no standard way to calculate an RP onset to this day (see Verbaarschot, Farquhar, & Haselager, 2015). Two ways were proposed by Libet, Gleason, et al. (1983, p. 632): (1) an RP_{MN} method (MN = *main negative shift*) and (2) $RP_{90\%}$ method. The RP_{MN} method is a straightforward method of eye-ball inspection of the waveform checked by a second independent investigator. The $RP_{90\%}$ is based on calculating the area under the RP waveform and places the onset to the point where the area (cumulating backwards from the EMG onset) reaches 90% of the total area (for details see Libet, Gleason, et al., 1983, pp. 632–633 and our adaptation of this method in chapter 5.2.6 of this text). Libet admits that the two methods of the onset assessment diverged significantly in some cases and that the RP_{MN} method was not always completely unambiguous (Libet, Gleason, et al., 1983, p. 634).

So far, I have mainly described the EEG findings in the M, W, P and Pv series. In the S series (skin stimulation at unknown times), no RP waveforms are expected (as no movement should be intended by the participants), and therefore the S series could be considered a "divergent validity check" for the RP findings. Indeed, the EEG recordings before the stimulus delivery appeared flat with only occasional irregular deflections (Libet et al., 1982, p. 330). Libet emphasizes the contrast of these results with the EEG recordings from the M and W series in the same sessions, suggesting that RPs are not present if no motor action is planned. Additionally, Libet also reports occurrence of a large P300 waveform following the stimulus delivery, arguing that this might be considered evidence for good attentiveness of the participants (Libet et al., 1982, pp. 330–331).

Finally, in the Sp series, in which a skin stimulus was delivered at pre-set times, the pre-event potentials varied from virtually none to small and brief negative shifts (Libet, Wright, et al., 1983, p. 370). Libet interprets this as evidence that the CNV-like negativities found in the P and Pv series cannot be explained by non-motor processes, which might be suspected by similarities of the P and Pv series with the S_1-S_2 arrangement typical for CNV studies (see above).

2.4.2 Results of introspective reports analyses

In this chapter, I will talk about the reports of subjective impressions of initiating a movement (M reports), intending or having an urge to move (W reports) and registering a skin stimulus (S reports). Values reported here are relative to the event onset (i.e. EMG activation or stimulus delivery); for example, if a W report is -200 ms, then the time of the intention to move was reported to occur 200 ms before the movement onset recorded by EMG. The overall Libet's results regarding the introspective reports are summarized in Tables 1 and 2 in Libet, Gleason, et al., (1983, p. 630–631).

Libet used the S reports to adjust the M and W reports with the rationale that a systematic bias in the S reports must be linked to some more general *subject's bias* present also while reporting the M and W time (Libet, Gleason, et al., 1983, p. 631). The grand-averaged S report (i.e. the mean S report across all participants and all sessions) was -47 ms. This means that the participants reported the impression of receiving a stimulus on average 47 ms before it was actually delivered. Mean M report was even more negative than mean S report; grand average of the M reports was -86 ms and if "corrected" for the subject's bias by using the S report, it remained negative to the EMG onset by approximately 40 ms (Libet, Gleason, et al., 1983, p. 631). Consistently with the expectations, the grand-averaged W time was found to be reported at -204 ms, that is before both the movement itself and the report of perceived movement onset (i.e. the M report). Libet does not explicitly state the "corrected" W time but based on the same logic used in the corrected M reports, we can assume that it is about -160 ms. The order of the values together with mean RP onsets are presented in Figure 4.
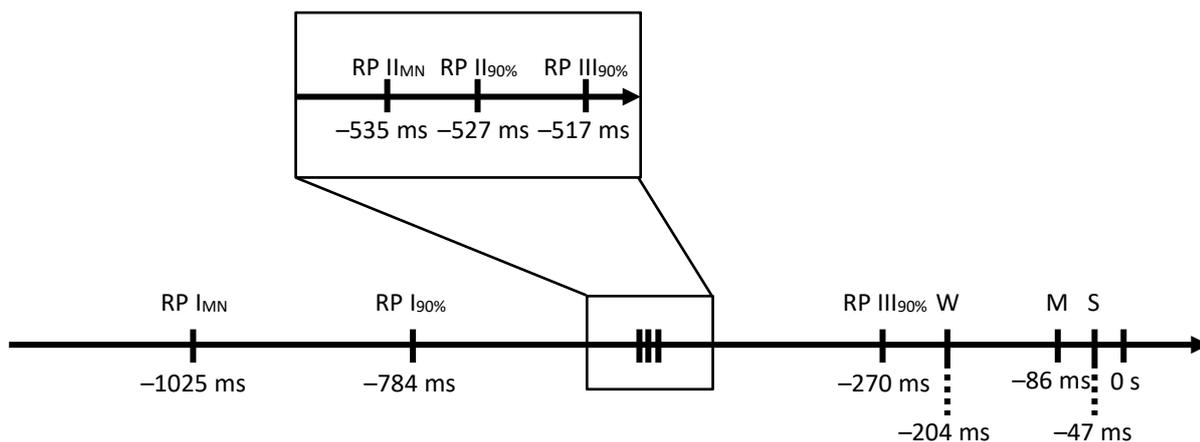
Figure 4: A timeline containing point estimates for (1) onsets of the main three RP types obtained by either the RP_{MN} and $RP_{90\%}$ methods and (2) introspective reports (M, W, S).

The key aim of Libet's experiment is to explore the temporal relation between neural processes in the brain and conscious intention to act (Libet, 1985, p. 530). It is clear from the presented results that the onset of RP of any kind generally tends to precede the W reports. However, it is difficult to assess statistical significance for such difference, given that standard deviation for the RP onsets cannot be calculated due to the necessity to average the EEG signal, hence losing the information on the RP onsets variance. This also means that it is impossible to make one-to-one comparison of each W report and the respective RP preceding it (Libet, Gleason, et al., 1983, p. 635). Nevertheless, Libet conjectures that W and RP should be positively correlated, and makes three points supporting this: *"(1) W and RP are features of the same underlying process, (2) observed differences between averaged RP and mean W were consistently large, and (3) there was a nearly complete absence of individual W values that were negative to the averaged RP"* (Libet, Gleason, et al., 1983, p. 635).

2.5 Libet's conclusions

So far, Libet's experiment may appear complicated and poorly aimed. However, in Libet (1985), a comprehensive framework for the previous three papers (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, Wright, et al., 1983) is offered, which I will summarize in this chapter. Libet states that the main objective for the whole experiment was to *"compare the time of onset of the conscious intention to act and the time of onset of associated cerebral processes"* (Libet, 1985, p. 530). To do that, he utilized the readiness potential recordings to

25

explore the cerebral processes and the rotating spot method to acquire the introspective reports, e.g. of the intention to act.

Regarding the RP recordings, Libet makes a few notes comparing his methodology to other studies interested in the RPs, such as Kornhuber's and Deecke's original research (1965). According to Libet (1985, p. 531), these studies have several downsides. First, the participants were required to perform an action repeatedly, inevitably leading to boredom and monotony. Second, the participants were asked to perform an act within a specific time window, possibly forcing them to perform it even when they did not wish to. Third, participants were not allowed to blink, again potentially forcing them to perform the act as soon as possible to relieve their eyes by blinking. Libet claims that these three limitations are not present in his experiment: the first problem is solved by asking the participants to report the M and W times after each movement, presumably breaking the emerging monotony; the second problem is eliminated by allowing the participants to perform the act whenever they feel like it; and the third problem is at least diminished by allowing the participants to blink during the task, with the condition that they then wait for at least one clock revolution.

Regarding the act itself, it might come into question why exactly Libet chose a flexion of a wrist or fingers. Libet, Gleason, et al. (1983, p. 640) point out that their design eliminates the factor of external values, beliefs and experiences, because a simple flexion of a muscle is an inconsequential action relatively free of influence from past experience. It should be noted here, that this inconsequentiality of actions might be considered an advantage (as Libet, Gleason, et al., 1983, suggest), but also as a flaw in the design, limiting its ecological validity (see Breitmeyer, 1985; Papanicolaou, 2017). While we could agree with Libet that a potential effect of past experience would pose a confound hard to control in an experimental setting, it is important to keep in mind that Libet's conclusions are relevant to isolated choices of *when or whether to act*, but only suggestive when studying the deliberation about *what to do* (see Libet, 1985, p. 536). Schlegel et al. (2015) describe this as the difference between *distal acts of willing* (long-term deliberation in a complex contextual environment) and *proximal acts of willing* (a simple act with virtually no deliberation, such as in Libet's study). We should add that there were voices, even in Libet's times, arguing that proximally deciding whether to act is a process incomparable to the distal deliberation process (Danto, 1985).

Importantly, Libet devoted a lot of attention to the question of validity of his methods, especially in the case of reporting the intention to move, which is crucial for his conclusions. He admits that he relies on the ability of his participants to associate a mental experience with the clock position (Libet, 1985, p. 534). The question is, is anyone capable of accurately assessing anything by this method? Libet (1985, pp. 534–535) presents several arguments for the clock's validity: (1) individuals were more or less precise in the S series, suggesting that such timing of an external event is reasonably accurate, (2) correction of the subjects' biases by their S reports magnified the difference between the reports and the RP onsets, not vice versa, (3) two purely mental events (the M and W times) were reported in consistently different times and (4) both modes of recall (absolute and order) converged to similar reports. Interestingly, Libet admits that it might be possible that the moment of the very first intention to move might not be retrospectively recallable; nevertheless, Libet states that *"such a hypothesis cannot be excluded since it is presently not experimentally testable"* (Libet, 1985, p. 535).

There are many possible objections to the validity of measuring the M and W times which Libet did not address here (many of which I present in chapters 3.1 and 3.2), but there is one which should be mentioned at this point. In 1998, Gilberto Gomes published a polemics with Libet's studies, generally challenging Libet's early backward referral hypothesis; however, he also makes a few points about the introspective reports Libet acquired in his 1980's experiments. He claims that in his view, there is no difference between experiencing an intention to move and experiencing the onset of the actual movement, ultimately suggesting that Libet might have simply induced an expectation in his participants that they should have some urge to move. Libet (2000) replied by two arguments: (1) participants themselves stated that they had no difficulties discerning between the M and W reports and (2) the quantitative difference between the M and W times was substantial with only a little variance. Obviously, this is not a satisfying answer, given that both can be true even when the participants' reports of the intention to move are artificially induced. Gomes (2002) noted that Libet (2000) did not in fact address his arguments. This issue remains open, even though it is crucial for the interpretations of Libet's findings. If Gomes is right, then the key element of Libet's interpretation, that is the difference between the W reports and the RP II onsets, is invalidated. I address this issue in detail in Study 1 (chapter 4), where our experiment regarding this particular problem is presented.

Libet (1985, pp. 535–536) also offers an interesting discussion of his RP recordings validity. He argues that it remains possible that some neural preparatory motor process starts in some other part of the brain than the supplementary motor area, but also notes that this would in no way invalidate his conclusions (since this would mean that the RP is actually a symptom of a late phase of the process, suggesting that the neural event starts even longer before the reported intention to act). Libet also discusses an alternative hypothesis about how RP is generated, which is of great interest today, because it foreshadowed the "stochastic accumulator hypothesis" proposed recently by Schurger, Sitt, & Dehaene (2012). In this hypothesis, Libet (1985, p. 535) states that it might be argued that RPs are generated even without any volitional process going on and instead arise spontaneously and repeatedly, while conscious control can be exerted to "activate" the covert motor preparation process underlain by currently ongoing RP. This topic is another crucial element of the interpretation of Libet's results, and I will therefore grant it appropriate attention in chapter 3.4.

Libet is mostly known among the laymen for showing that voluntary process starts unconsciously, which might obscure the fact that Libet in fact **opposed the idea of no free will**. In section 4 of his 1985 paper (pp. 536–538), he argues that conscious functions may exert control over the unconsciously initiated movement in the form of a conscious veto (meaning that an individual can voluntarily block the arising intention to act) or a conscious trigger (meaning that an individual can voluntarily impel the intention to act). While the conscious trigger hypothesis was not directly tested by Libet (to my knowledge), Libet (1985, p. 536) states two arguments for the existence of a conscious veto. First, participants in the original experiment (Libet et al., 1982; Libet, Gleason, et al., 1983) occasionally reported having an urge to act which was later "aborted". Second, the study reported in Libet, Wright, et al. (1983) demonstrated that in the Pv series, an RP was generated even when the movement was successfully blocked (which was validated by no EMG recording). It should be noted that the idea that the "conscious individual" can exert some control over the "unconscious brain process" is somewhat dualistic and hence hardly defendable. As one might expect, more recent study, authored by Brass & Haggard (2007), already suggested a candidate brain area (specifically fronto-medial cortex) for the veto exertion.

3 Critique and discussions of Libet's experiment

Because Libet's experiment is so impactful on how we relate to the questions of free will and sense of agency, it is being intensely discussed and its features thoroughly examined even to this day. In June 2018, I attended the 22nd annual meeting of the Association for the Scientific Study of Consciousness (ASSC) and noted that a substantial proportion of all contributions presented at the meeting pertained either Libet's experiment itself or closely related topics of agency and deliberation (see ASSC, 2019). Nevertheless, the amount of scientific attention paid to Libet's experiment is not a recent phenomenon. In this chapter, I would like to present some of the most important discussions from throughout the last three decades. While presenting some of the discussions, I will also present my opinion on them. My reason for discussing these myself is that I would like to offer some objections to their interpretations. However, I would also like to state that **I do not intent to dismiss these studies' validity**; rather I would like to propose constructive counterarguments in accordance with the principles of scientific discussion.

3.1 Discussions of the intention to move (W reports)

Regarding the introspective awareness of the intention to act, I have already addressed the argument between Gilberto Gomes (1998, 2002) and Benjamin Libet (2000) regarding whether the M and W times are discernible. Gomes based his objection partially on Breitmeyer's (1985) claim that he himself and some of his colleagues tried to perform an action just as Libet required his participants and noted that they were unaware of any preceding intention to act while doing so. Breitmeyer states that *"by requiring subjects to attend to awareness of intent, Libet may have imposed intention artificially and in a way that is not comparable with more ecologically and existentially valid voluntary and intentional acts"* (Breitmeyer, 1985, pp. 539–640). This idea was further advocated by Pockett & Purdy (2011), who pointed out that the W reports might in fact be inferred rather than perceived. These claims are substantially supported by empirical findings. Banks & Isham (2009) conducted two experiments, both of which demonstrate that the W reports might be manipulated by providing a delayed auditory or visual feedback. Similarly, Lau, Rogers, & Passingham (2007) showed that by delivering a pulse via transcranial magnetic stimulation (TMS) they were able to manipulate the W reports in a similar fashion to Bank's and Isham's trials. These two findings suggest that the W reports are in fact ex post facto inferences, since

they can be retrospectively biased by an experimental intervention. In chapter 4, I report our own contribution to this discussion, showing that the W reports might be inferred based on participant's experience with reporting the onset of the actual movement.

Despite—or maybe because of—these controversies, a large amount of research was devoted to the intention to act since Libet's original publications. Matsuhashi & Hallett (2008) found that using different methodology, they were able to obtain an estimate of the mean time of the intention to act equal to 1.42 s before the movement onset. Their inventive design was based on the following task: the participant was asked to perform a movement in a similar manner as in Libet's original study, while a tone was presented at irregular times. Participants were instructed that if the tone was presented while they were thinking about the movement, they should abort it. If the tone was presented before any thought of the movement or when the movement was already being performed, they were asked to simply ignore the tone. This way, the authors were able to analyse the distribution of the tones relative to the movement onset, expecting a disruption in otherwise uniform distribution in the period when the intention was being processed in participants' brains. The results fitted the hypothesis, suggesting that an intention to move can be traced back even to almost one and half a second before the movement onset. Authors propose several interpretations to explain why their results do not fit Libet's conclusions, one of which seems particularly intriguing—they use analogy with mind-wandering studies, which show that participants spontaneously realized that they are mind-wandering ("self-caught") on relatively rare occasions, but if interrupted and asked directly ("probe-caught"), they reported having mind-wandered significantly more often (Smallwood & Schooler, 2006). Similar mechanism might apply in Matsuhashi's & Hallett's (2008) results, since asking the participants to abort their movement when a tone is presented is equivalent to interrupting a mind-wandering person, while asking the participants about the intention retrospectively—as Libet did—equals relying on self-reported mind-wandering. Of course, it might be possible that the results only reflect participants' effort to comply with the instructions and not perform any action closely following the tone (as this would violate the explicit instruction). Matsuhashi & Hallett address this objection by stating that if this was the case, it would be unlikely to obtain similar results across participants and to arrive at conclusions compatible with previous research. The authors further add that they specifically told the participants to avoid movement within some period after the tone, if the participants themselves reported such

strategy. I would like to advise caution here by pointing out that (1) there was a large variance in identified times of intention between participants (see Table I in Matsuhashi & Hallett, 2008) and that (2) Matsuhashi's & Hallett's ICC (intraclass correlation coefficient, see for example Weir, 2005) seemed to reflect only „slight" level of test-retest reliability (see Shrout, 1998) for the W reports. This amount of both inter- and intraindividual variability raises the question whether this method of estimating the time of intention to move is more reliable than Libet's clock, given that the results are remarkably different. Nevertheless, a recent study showed that the probe method suggested by Matsuhashi & Hallett (2008) might prove more valid than the requirement for direct introspection present in the Libet's clock method, because the introspection seems to notably alter brain signals, including the readiness potential characteristics (Verbaarschot, Haselager, & Farquhar, 2019).

Another important suggestion regarding the reports of intention to move was made by Schlegel et al. (2015), who showed that an endogenous movement, uncued by any external stimuli, can be made without any experience of conscious intention to move, if the movement is induced within a posthypnotic suggestion. Furthermore, these authors demonstrated that the readiness potential waveform did not differ between a consciously willed and posthypnotically suggested movements. These results imply that the RP might be connected to processes unrelated to the process generating conscious intention to act. Nevertheless, authors emphasize that studies like this one should be considered with caution, given that hypnosis is still imperfectly understood phenomenon and that the sample appropriate for analysis was reduced to only four participants due to a stringent requirement for complete posthypnotic amnesia.

Studying the intention to act from a neuroscientific perspective, Lau, Rogers, Haggard, & Passingham (2004) used fMRI scanning of participants who were asked to report the time of their intention to act (i.e. the W times). They found that compared to the control condition—in which the participants were required to report the time of the movement onset (i.e. M times)—an enhancement of activity in the pre-supplementary motor area (pre-SMA) was found in the W condition. The data were further re-analysed in a follow-up paper (Lau, Rogers, & Passingham, 2006), which found that there is a negative correlation between the W reports and pre-SMA activity enhancement. This means that the attentional process biases the W reports; authors even arrive at the conclusion that if we could eliminate this attentional bias, then the W reports would not be detected at about -200 ms, but at -120 ms, which would

ultimately discard any suggestions regarding the conscious veto, because there would simply be not enough time for the veto to intervene.

Similarly, Desmurget et al. (2009) made interesting findings regarding localization of the intention to act in the brain. By stimulating 57 brain areas in frontal, parietal and temporal lobe with direct electrical stimulation (DES) they found, among others, that stimulating 16% of these areas produced a feeling of a movement or intention to move. This phenomenon was especially pronounced in BA 39 and BA 40[1] in the posterior parietal cortex, where the stimulation elicited strong urge to move with no overt EMG activation. Two of their seven participants even reported having moved, again with no EMG activation detected. Conversely, stimulating dorsal premotor cortex (BA 6) elicited a movement without any accompanying awareness of it. In the context of Libet's experiment, these results suggest that not only the motor preparation process, but also generation of a conscious urge to move has neural correlate in the brain, which ultimately questions Libet's general idea of "conscious control" over an "unconscious brain".

Desmurget et al.'s findings are further supported by preceding research by Sirigu et al. (2004), who found that patients with selective parietal lesion tend to provide extraordinarily late W reports. Sirigu and her team interpret this as a sign that parietal cortex participates on connecting the motor process to higher mental and conscious processes but is not in itself responsible for the movement execution, since the parietal patients were otherwise able to perform a movement normally. Furthermore, the authors reported that the parietal patients did not differ from the control group in the M reports or auditory S reports.

At the first glance, findings of Desmurget et al. (2009) and Sirigu et al. (2004) seem to contradict Lau et al.'s (2004, 2006) findings—Lau et al. identified the activity in pre-SMA, while Desmurget and Sirigu talk mostly about the parietal cortex. However, Desmurget et al.'s method was based on exploring brain areas responsible for elicitation of **the sensation of urge or intention** to move and, similarly, Sirigu et al. studied abnormalities in **the reports of the intention** to act. Whereas Lau et al. were interested in the influence of shifting one's **attention** from the movement itself to the intention to move. Therefore, both groups of findings are informative and compatible—Lau et al. (2004) suggest that requiring the

---

[1] BA stands for Brodmann area. For an overview of the Brodmann brain map, see for example Trans Cranial Technologies (2012).

participant to pay attention to the intention to act changes the neural processes preceding a voluntary movement in pre-SMA, while Sirigu et al. (2004) and Desmurget et al. (2009) seem to agree on the fact that the intention to act itself might arise in the posterior parietal cortex.

However, reality might be more complicated than these findings suggest. Guggisberg, Dalal, Schnider, & Nagarajan (2011) studied decisions of which hand to move after an imperative signal. There were three basic introspective conditions, based on which mental event the participant should attend (either the signal, the decision or the movement itself). Using MEG (magnetoencephalography)—and thanks to some important features of the design—the authors were able to identify neural correlates specific to the process of introspection of an event, which was, however, **not coding the event itself**. Most notably, angular gyrus was found to be related to the introspection of the decision of the hand, while its activity was unrelated to previously making the decision. This has several interesting implications—it suggests that there are specific neural networks responsible for introspection of different mental events and that the neural activity observed by Libet et al. might be, to significant extent, influenced by the need to introspect a mental event. This last suggestion might be challenged by pointing out that in the S series, no negativity was found, even though the stimulus timing was introspected, but overall the conclusions seem meaningful; they also fit previous Desmurget et al.'s (2009) findings about BA 39, which happens to encompass the angular gyrus (Trans Cranial Technologies, 2012).

Miller et al. (2010) studied the decision time in slightly modified sense. The authors were interested in decisions whether to act in discriminatory reaction time (RT) tasks, in which participants are asked to respond to certain stimuli but inhibit the response to different stimuli. Based on results of their four experiments, in which they presented the RT task in several different variants, they reported that the decision times given by their participants were notably inaccurate, relatively easily manipulated by experimental factors and in some cases even implausibly early or late. Although the experiment's conclusions are informative and interesting, I would like to state my doubt whether it is possible to compare the W times in Libet's sense (as times of intention to act) and the reported times of decision whether to act in discriminatory tasks. Firstly, I would argue that Libet explicitly stated that he studied voluntary movement, which he defines as arising *"endogenously, not in direct response to an external stimulus or cue"* (Libet, 1985. p. 529), so Libet's first objection to Miller et al.'s

findings would probably be that reaction time tasks are not endogenous and therefore cannot be driven by the same neural mechanisms. Secondly, I am not convinced that the decision whether a stimulus was a "go" or "no-go" is conscious in the moment it occurs. As with many reflexive acts, it may very well be that the decision whether to respond to a stimulus is processed unconsciously, while the conscious realization of such decision may be generated retrospectively. Most importantly, regardless of whether this conjecture of mine is true or not, it does not imply that the same must apply for the W time in Libet's sense. This question might be worth further investigation.

Overall, the W reports seem to be mostly unreliable, even though in some context the data make sense—for example when correlated with the amount of neural activation, as stated in Lau et al., 2004. I also discuss the validity of using Libet's clock for measuring the W times in chapter 4. Nevertheless, however unreliable the **temporal** judgements of the intention to act might be, the subjective feeling of intention or urge to act should not be dismissed **as such**, because it is an experience predictively informing us about our urges, albeit temporally inaccurately.

3.2 Discussions of the reports of the movement onset (M reports)

Some researchers studied the introspective process by investigating the reports of the movement onset (i.e. M reports in our terminology). Pockett & Miller (2007) studied specifically the M reports made in the absolute (A) mode of recall. The authors manipulated seven dichotomous characteristics of the clock to explore their effects (see Table 1). They tested 64 factor combinations and evaluated 7 main effects and 21 two-factor interactions. Results showed three significant main effects for factors *Start/End Movt*, *Dark/Light Spot* and *Centre/Follow*. *Start/End Movt* effect suggests that when the participants were asked to report the time of the keypress, they reported later time by 32.4 ms compared to when they were asked to report the start of their finger movement. *Dark/Light Spot* effect showed later M reports by 14.8 ms when the spot was lime green in comparison to a dark blue spot. *Centre/Follow* effect produced earlier M reports by 19.3 ms when the participant was asked to follow the dot around the clock compared to when he or she was asked to fixate the centre of the clock. Furthermore, an interaction of *Start/End Movt* and *Little/Big Spot* was found. After adjusting for multiple comparisons, only the *Start/End Movt* effect remained significant, which was further confirmed in a follow-up experiment. This was expected, since authors point out that if the start of the movement and the keypress, which occur in

objectively different times, could not be distinguished, then the clock method should not be considered accurate in any sense. Authors derive several conclusions. First, they note that Libet's clock method is surprisingly robust, at least for investigating the M reports. Second, they make a recommendation to obtain M reports with low variability: ask participants to make a clear decision to move and use a little spot and high clock speed; on the other hand, the interindividual variability remains high regardless of these factors. Third, the clock size, the spot size and spot speed were not found to be significant factors influencing the mean reports, which suggests that results arrived at in different laboratories are comparable.

Factor	Condition 1	Condition 2
X1 Urge/Decision	Allow urge to appear spontaneously	Make definite decision to ACT NOW
X2 Start/End Movt	Report start of finger movement	Report time keypress complete
X3 Dark/Light Spot	Dark blue spot	Lime green spot
X4 Little/Big Circle	Circle radius 1.3 cm	Circle radius 2.6 cm
X5 Centre/Follow	Fix gaze on centre of circle	Follow spot round with eyes
X6 Little/Big Spot	Spot diameter 2 mm	Spot diameter 3.5 mm
X7 Slow/Fast Rotation	Rotation speed 3.19 cm/s (Period 2.56 s for small circle, 5.12 for large circle)	Rotation speed 6.38 cm/s (Period 1.28 s for small circle, 2.56 s for large circle)

Table 1: Clock characteristics studied in Pockett & Miller (2007). The table is directly copied from Table 1 in the original paper (Pockett & Miller, 2007, p. 243) with the permission of the first author.

Lau et al. (2006) studied the M reports using an fMRI scanner in an experimental factorial 2×2 design. One factor was **modality** of the report—either a movement onset (*action condition*) or auditory stimulus delivery (*auditory condition*); the other factor was **timing**—in the *timing condition*, the participant observed the clock just as in Libet's original experiment, whereas in the *non-timing condition*, the spot was displayed only during the first revolution, then disappeared and appeared shortly only in the moment of the event, i.e. the keypress or stimulus delivery. They found that the *timing* factor influenced activity in brain areas connected to visual processing, while the *modality* factor influenced different brain areas based on the modality. The auditory condition increased activity in "*right primary*

motor cortex, the supplementary motor area and subjacent cingulate cortex, left supramarginal gyrus and right ventral thalamus" (Lau et al., 2006, p. 7268); the action condition increased the activity in right superior temporal sulcus. Most importantly, however, a negative correlation was found between the M reports and brain activity enhancement in cingulate motor area (CMA) in the timing condition compared to the non-timing condition. This suggests that while previous research shows specific connection between the W reports and pre-SMA (Lau et al., 2004), equivalent neural basis for the M reports seems to be CMA.

Obhi, Planetta, & Scantlebury (2009) investigated the M reports in surprisingly simple, but largely informative way. They based their hypothesis on the question whether the anticipatory M judgements (by anticipatory they mean early) are caused by pre-movement processes (efferent account) or by processes which follow the movement onset (reafferent account). They observed differences in the M reports when the movement was (a) made by a finger or a toe (factor *effector*) and when it was (b) made actively (initiated by the participant) or passively (experimenter pushed participant's finger or toe; factor *source of movement*). The results showed significant main effect of the *effector* with less anticipatory (i.e. later) M reports for the toe movement, as well as significant main effect of the *source of movement* exhibiting less anticipatory M reports for a passive movement. No significant interaction was found. The difference in subjective timing between active and passive movement supports the efferent account of the bias, essentially suggesting that Libet's classical M reports are early because they reflect processes ongoing in the brain when the movement is being prepared, which does not happen when the movement is made passively. However, significantly later M reports for the toe movement compared to the finger movement suggest that the M reports must be additionally influenced by reafferent factors, taking longer time when carrying the information from a toe than from a finger.

Overall, it seems that the M reports are much easier to study, as they are subjected to lower amount of bias than the W reports, while this bias seems to be influenced by relatively simply identifiable factors (such as the efferent and reafferent factors). In conclusion, unlike those of the W reports, the M reports' reliability and validity is well documented.

3.3 Discussions of the reports of skin stimulus (S reports)

In this chapter I will mention a few contributions to discussion about the S reports. Gomes (1998) pointed out that the S reports in Libet's original study might have been influenced by

the feedback which Libet and his team continually provided to the participants within both the re-training S series at the beginning of each session and the regular S series (see chapter 2.3.3 of this text and Libet, Gleason, et al., 1983, pp. 627–628). Libet believed that this introspection accuracy training was crucial, because a systematic bias in the S series would also apply to other tasks, such as the W series (hence the suggested correction for the M and W reports, see Libet, Gleason, et al., 1983, pp. 630–631). However, Gomes (1998) points out that this feedback training performed on a regular basis throughout the experiment might lead to variable results across the sessions. Additionally, it seems poorly justified to use the training based on the skin stimulation specifically, because Danquah, Farrell, & O'Boyle (2008) showed that the S reports differ for tactile, visual and auditive modality—this implies that the reports provided after an auditory training would presumably differ from those provided after a tactile training. These findings complement Breitmeyer's (1985) objection that the idea of "correcting" the M and W reports by the mean S reports for given participant is misleading, given that there are no guarantees that sensory and introspective timings are equal.

Besides this issue, however, I am currently unaware of any substantial bias specific to the S reports currently discussed in the literature. However, the literature suggests several biases, which might presumably influence participants' reports regardless of whether they are M, W or S. One of these general biases is called the **flash-lag effect**. The flash-lag effect was first hinted by MacKay (1958), only to be "rediscovered" in 1994 by Romi Nijhawan, who was interested in how it is possible that we can perform heavily time-precise actions, such as catching a ball, given that our perceptual system processes the inputs relatively slowly. Nijhawan hypothesized that this is thanks to the ability of our perception the extrapolate current position of a moving object and estimate its position in the next instance. He supported this hypothesis by demonstrating a phenomenon which he called flash-lag effect. Nijhawan's version of the experiment utilized a screen on which two lines of the same orientations but different lengths were rotating at 30 RPM. The shorter line was displayed continuously, while the longer one was flashing on the screen with a duration of only 5 ms. Nijhawan's participants observed that the lines did not appear to be aligned—the shorter line appeared to be slightly ahead of the longer one at the instant of each flash. Nijhawan interpreted this finding as an evidence for the extrapolative perceptual process applied to the continuously displayed line, but not to the strobed line. This interpretation was subjected to

some critique (see for example Baldo, Kihara, & Namba, 2002), but the phenomenon's existence itself is very important for Libet's experiment. It suggests that the position of the dot sliding along the clock's circumference might be perceived to be slightly ahead of its actual position. Therefore, this effect might influence any reports in any mode of recall made using the Libet's clock, as pointed out by many researchers (Klein, 2002; Kawohl & Habermeyer, 2007; Matsuhashi & Hallett, 2008). Nevertheless, it should be added that some argue that the flash-lag effect is in fact negligible in the case of the Libet's clock (Pockett & Miller, 2007).

Another potential general bias is so called **prior entry effect**. This phenomenon was already mentioned by Titchener (1908, in Sternberg & Knoll, 1973). Prior entry effect basically means that "*events in an attended stream appear to occur earlier than simultaneous events in an unattended stream*" (Haggard & Libet, 2001, p. 49). In context of Libet's experiment, the prior entry effect could potentially bias any introspective reports because they refer to—from the nature of the experimental task—attended mental experiences. Breitmeyer (2002) notes that the role of prior entry effect was mentioned, but not sufficiently addressed by Libet (1985). On the other hand, Haggard & Libet (2001) stated that, with respect to previous experimental findings, the prior entry effect would be too small in magnitude to significantly alter Libet's original results, which opinion was later repeated by some researchers (Matsuhashi & Hallett, 2008), but also questioned by others (Papanicolaou, 2017). Interestingly, Cairney (1975) stated—and empirically supported—a notion that the prior entry hypothesis might be applicable to the rotating spot method to significantly limited extent, maybe even not at all.

3.4 Discussions of the readiness potential (RP)

Debates concerned with readiness potential seem to currently dominate the Libetian discussions. They include questions like whether the RP is truly movement-specific, what its origin is, whether there are better ways to monitor movement-related preparatory activity etc. I will introduce here some of the most important findings in this regard.

In 1990, Keller & Heckhausen conducted three experiments to expand on Libet's results. In the Experiment 1, the participants were asked to perform a mental counting task, but if they moved, they were asked whether they were aware of the movement and whether the movement was pre-planned. This way, the authors collected data on motor preparation process in unconsciously initiated movements. The Experiment 2 was essentially replication

of the W series in the A mode of recall. Results of these two experiments showed similar RPs to what Libet had found. Authors note that the RPs preceding an unconscious movement were smaller in amplitude than those preceding a voluntary movement. Mean difference between the RP onsets and the W reports was also similar to Libet's results; interestingly, this difference was remarkably consistent in seven participants out of eight, which is in notable contrast with later studies exhibiting large interindividual differences (e.g. our Study 2). In the Experiment 3, the participants were asked to sit calmly, relax and introspectively observe their right and left arm; if a movement occurred, a procedure similar to the Experiment 1 was carried out. Results showed that in this condition, unconscious movements were much rarer than in Experiment 1. When asked about their introspective experience, the participants provided two kinds of reports: either that they were unaware of having moved or that they experienced a suddenly arisen feeling of "wanting to move". Readiness potentials obtained in Experiment 3 did not differ from those found in Experiments 1 and 2. Overall, Keller and Heckhausen corroborated some of Libet's conclusions.

More critically, Haggard & Eimer (1999) provided an argument that RP seems to be causally unrelated to the W reports, whereas so called lateralized readiness potential (LRP) might be. The LRP can be obtained if the experiment contains movements with both the left and right hand. It is calculated as a "double difference", that is as a difference between the activity over the left and right motor cortex (labelled usually as C_3' and C_4') when the movement is made with the right hand, subtracted from the same brain activity difference when the movement is made with the left hand (Eimer, 1998). The calculation can be summarized in the following formula (Eimer, 1998, p. 148):

$$LRP = (C_3' - C_4')_{\text{left hand}} - (C_3' - C_4')_{\text{right hand}}$$

Haggard & Eimer (1999) noted that according to arguments of John Stuart Mill, a covariation between two phenomena is a characteristic aspect of causality (in the original: "*Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation.*", Mill, 1882, p. 287). In their experiment, Haggard & Eimer did not find any covariation between the RP and the W reports, which according to Mill's logic suggests that there might be no causal relationship between the two. On the other hand, this expected relation was found between the LRP and the W reports, suggesting that the LRP is more tightly linked to the process which individuals refer to when reporting the W

times. The authors concluded that the M and W reports are connected to specific movement preparation rather than to non-specific preparatory mechanisms. Haggard elaborated on the findings in 2001 in a joint publication with Libet, who presented his counterarguments (Haggard & Libet, 2001). Libet argued that RPs found in Haggard & Eimer (1999) had notably early onsets, which according to him suggests that the participants predominantly pre-planned the movement. He further questions Haggard's and Eimer's methodology for the RP onset estimation and points out that the W reports found by Haggard and Eimer are different from what he has previously found himself. Overall, Libet argued that Haggard's interpretations are speculative and that the issue is still open to investigation.

Despite this disagreement, LRPs found their firm place in the libetian discussions. Trevena & Miller (2002) conducted a variant of Libet's experiment in which they studied the order of the RPs, LRPs and W reports obtained in the O mode of recall. They arrived at the conclusion that the hand-specific LRP mostly begins after the W report. This has an interesting implication that the RP may in fact reflect only general non-specific processes, which are on their own insufficient to elicit an action. The LRP, on the other hand, might be connected to the motor preparation more tightly and therefore should be more important for interpretations of Libet's conclusions. Because the LRPs were found to predominantly follow the W reports, a possibility arises that Libet's experiment would not bring that surprising results if LRPs were investigated instead of the RPs. Interestingly, the authors tested a hypothetical explanation that the early RP onsets found in Libet's original study might be caused by so called **smearing artefact**. The smearing artefact is a phenomenon when an onset of averaged waveform component tends to shift in time due to extremely early onsets of a few individual waveforms contributing to the average. Nevertheless, within their Experiment 1, Trevena & Miller (2002) concluded that the smearing artefact cannot account for the early RP onset and that it seems that the readiness potential does indeed start before the mean W report.

In 2010, Trevena & Miller expanded on this issue when they conducted two experiments. In Experiment 1, the participants waited for a tone and were instructed that when the tone sounds, they should decide whether or not they will move ("*sometimes-move*" condition) or move right away ("*always-move*" condition). The authors found longer reaction times (RTs) in the sometimes-move condition, which according to the authors suggests that it takes some time to make a spontaneous decision whether to act. More importantly, Trevena & Miller

also found EEG negativity preceding the tone in all conditions (that is "*always-move*", "*sometimes-move*" if the movement occurred and "*sometimes-move*" if the movement did not occur). This is somewhat surprising, since the tone could not be expected. The authors link this pre-tone negativity to participant's engagement in the task. On the other hand, LRPs found in the experiment corresponded better with the expectations. Not only that the LRPs followed the tone, they also only occurred if the participant performed a movement. This further suggest that LRPs are linked to the movement itself more tightly that general RPs. In the Experiment 2, authors added the instruction that after the tone, the participants were supposed to decide which hand they will perform the movement with (regardless of whether it was in the "*always-move*" or "*sometimes-move*" condition). Again, results showed that LRP starts after the tone, suggesting the participants indeed decided which hand to move after the tone.

Even though this study corresponds with previous findings regarding the LRP, I found a few points problematic and maybe needing further explanation. First, the pre-tone negativity found by Trevena & Miller (2010) does not resemble RP in any way, except that it rises steadily. The amplitude increase seems to be only 1 µV per second and it does not have the typical ramp-like or dome-like shape. Therefore, comparing it to the RP does not seem intuitive to me. Second, the paper implies that the pre-tone negativity must have been present constantly during the trial, because the tone was presented at a random time unknown to the participant. This is another reason why it should not be compared to the RP—in my opinion—since it seems more like a continuous baseline drift (which Libet accounted for, see the dashed lines in Figures 1, 2, 3, 4 and 5 in Libet et al., 1982), not an event-related potential. Furthermore, for the same reason, it seems unjustified to set any form of baseline, as the authors did, because since the negativity must have been rising steadily and the tone could not be expected at any certain time, it makes no sense to assume a baseline when the voltage should not change. Third, the authors seem to consider the decision *whether to move* (as in their study) to be similar to the decision *when to move* (as in Libet's study). In fact, this is the same issue I pointed out when discussing Miller et al. (2010) in chapter 3.1—I do not understand what rationale there is for the assumption that the decision *whether to move* is the same as the decision *when to move*. Fourth, when the authors discuss how their experiment's results relate to Libet's original study (given that they study a decision made after an imperative signal), they mention an explanation that the participants might have

adopted a strategy of always deciding to move and only sometimes making a subsequent decision not to. This explanation seems to me to be the most realistic, but authors dismiss it stating that *"a negativity difference between move and no-move trials would indicate that the veto decision was not influenced by the level of prior response preparation"* (Trevena & Miller, 2010, p. 455) and that the participants were not particularly motivated to perform fast responses. It is not clear to me why the veto should be influenced by the pre-tone preparatory activity—that is why cannot the participants decide to veto a movement after the signal— and why the participants should not be motivated to adopt this strategy, even when there is no incentive to react quickly.

Several studies employed advanced imaging techniques and supported Libet's conclusions. For instance, Fried, Mukamel, & Kreiman (2011) demonstrated that results similar to Libet et al.'s can be obtained even on a single-neuron level. Not only that, the authors also reported having been able to predict the decision to move on a single trial basis. Perhaps the most impactful publication within the post-Libet discussions, however, is a 2008 paper by Soon, Brass, Heinze, & Haynes. The authors based their study on Libet's findings regarding the RP but argued that the RP onset might in fact only reflect a late phase of the causal chain leading to a movement. Therefore, they studied whether there might be a specific brain activity influencing outcome of an action which can be traced even further back in time. Their participants conducted spontaneous movements with either left or right hand while their brains were scanned using an fMRI machine (functional magnetic resonance imaging). They reported time of the intention to move by selecting a letter they saw when the intention occurred; the letters were presented in a stream by 500 ms. This might seem to be inadequately rough, since this can provide maximal temporal resolution of 0.5 s. However, the method of fMRI scanning used by the authors was so called Blood-Oxygen-Level-Dependent imaging (BOLD, see e.g. Forster et al., 1998) which is known for its low temporal resolution (i.e. it is not as time-precise as EEG) and so, refined introspective data in orders of milliseconds would not be meaningful. The results showed that during the execution phase of the movement, an increase in activity was found in the primary motor cortex and supplementary motor area (SMA). More importantly, authors also found activity patterns coding which hand the participant will move in the frontopolar cortex (BA10) and medial parietal cortex (specifically, precuneus and posterior cingulate gyrus). In the frontopolar cortex, the authors reported identifiable patterns occurring already 7 s before the

movement occurred. Given that BOLD has a significant latency, it can be estimated, authors argue, that the activity is in fact present even 10 s before the movement. The authors propose an interpretation that the plan of which hand to move originates in the frontopolar cortex and is stored in the parietal cortex, until it becomes conscious. The authors conclude that *"this suggests that when the subject's decision reached awareness it had been influenced by unconscious brain activity for up to 10 s"* (Soon et al., 2008, p. 545). The impact of this paper is notable in subsequent publications, such as one titled *Mental Causation and Free Will after Libet and Soon* (Batthyany, 2009); I should concede that this publication advises caution when interpreting Libet's and Soon's conclusions, but the importance of Soon's findings is clearly articulated.

Since 2011, several studies investigated whether RP can be induced not by the action, but by the experimental situation itself. As already stated in chapter 3.1, Guggisberg et al. (2011) found that introspective effort might alter the neural activity preceding a voluntary movement. Antonietti (2011) points out the amount of "multitasking" in Guggisberg et al.'s experiment and notes that the metacognitive process is in fact hierarchically layered; therefore, we should be cautious when interpreting a neural activity specifically as a sign of preparatory motor process. The bias caused by multitasking in Libet-style experiments is shown by Miller, Shepherdson, & Trevena (2011) who demonstrated that (1) movement-preceding negativity is steeper when the clock-task is presented compared to when it is not and that (2) similar negativity can be observed when the clock is present, even though no movement is required. As I previously noted on other Miller's experiments, the authors claim—somewhat controversially in my opinion—that the overall increase in negativity throughout a trial is the cause for the emergence of an RP. I have already argued that these steady baseline shifts should not be directly related to readiness potentials, because they do not explain their morphology, and I have also noted that Libet did in fact account for these shifts, as can be seen in the Figures 1, 2, 3, 4 and 5 in Libet et al. (1982). Nevertheless, Miller et al.'s (2011) finding remains interesting: although it can be argued that the negative shifts should not be considered readiness potentials, it still shows that the mere presence of the clock can significantly alter the recorded EEG activity. Similarly, Alexander et al. (2016) showed that a waveform similar to RP can be detected in experimental situation with a purely mental decision, suggesting that the typical RP might reflect general anticipatory process rather than motor-related activity. Another cognitive factor potentially influencing the RP

might be internal timing, as suggested by Verleger, Haake, Baur, & Śmigasiewicz (2016). These findings overall suggest that the connection between readiness potential and pre-motor cognitive processes might be rather indirect.

One of the most influential papers on the nature of the RP is the 2012 study by Schurger, Sitt, & Dehaene, which is currently being intensively discussed (for example, see Schultze-Kraft et al., 2016; Alexander et al., 2016; Khalighinejad, Schurger, Desantis, Zmigrod, & Haggard, 2018) and sometimes referred to as an explanatory framework for RP generation (e.g. Schlegel et al., 2015). Schurger and his colleagues suggested that RP-like negativities might occur as spontaneous stochastic fluctuations in EEG voltage, regardless of whether a movement occurs. These fluctuations per se cannot be detected, because to detect RP-like waveforms the recording needs to be averaged after being time-locked on a certain event (which is not present in spontaneous EEG recording). In the light of this argument, Kornhuber & Deecke (1965), Libet et al. (1982) and many other researchers might have created a backward selection bias when calculating averages only from the epochs in which a movement occurred. It can be assumed that if these spontaneous stochastic fluctuations exist, then after accumulating enough of the physiological noise (to the extent that it exceeds a certain threshold), an overt movement can be elicited, which is then experienced as self-initiated and spontaneous act. Schurger et al. (2012) supported this hypothesis by an experiment with a design called *Libetus interruptus*. Participant's task was to watch Libet's clock, relax and make a spontaneous button press at any time. Nevertheless, in some cases the trial was interrupted by an audible click, which prompted the participant to conduct the movement immediately. The most important finding was that the third of the fastest reactions to the interrupting click was preceded by a long slow negative potential, while the third of the slowest reactions was not. The preceding negativity cannot be considered a specific movement preparation because the clicks were presented at unpredictable times. The fact that the fastest reactions were preceded by an RP-like negativity suggests that when the participant was more prepared to make the movement, the stochastic fluctuations were closer to the threshold than in the trials in which the participant was less prepared (i.e. had slower reaction times). This has significant implications for how we operationalize the onset of neural motor activity. While Libet et al. (1982) assumed that the neural activity leading to the movement begins when the RP deflects from the baseline, Schurger et al. (2012) suggest that this deflection does not have a meaningful interpretation and that the *"neural decision*

to act now" (Schurger et al., 2012, p. E2905) actually occurs when the negativity exceeds an activating threshold. This threshold crossing must logically occur later than the voltage deflections analysed by Libet; therefore, it might be even possible that when reporting the W times, the participants might report the instances of the threshold crossing, which would explain the discrepancy between the RP onsets and the W times reported in Libet, Gleason, et al. (1983).

Schurger's hypothesis was supported by several studies. For instance, Jo, Hinterberger, Wittmann, Borghardt, & Schmidt (2013) studied so called slow cortical potentials (SCPs), which are slow brain potentials with frequency lower than 1 Hz, which are presumably constantly present in EEG recordings (see Schmidt, Jo, Wittmann, & Hinterberger, 2016, p. 642). Jo et al. (2013) showed that voluntary movement is more likely to be preceded by a negative slope of an SCP than by a positive slope, compared to 50:50 ratio of positive and negative shifts in a control condition with no spontaneous movement. This suggests that the predominance of negative SCP slopes might be at least partially responsible for the averaged RP waveform. Authors note that this is compatible with Schurger et al.'s (2012) idea about spontaneous fluctuations provoking a decision to move. However, Schmidt et al. (2016) point out that their hypothesis (which they call SCP sampling hypothesis) differs from the Schurger's hypothesis in one important point: while Schurger's model presumes existence of a threshold which the accumulated fluctuation must cross to elicit a decision to move, the SCP sampling hypothesis does not. Recently, a study by Khalighinejad et al. (2018) showed a phenomenon which cannot be explained by the stochastic fluctuations account. The task used in this experiment was to wait for a set of chaotically moving dots to start moving coherently and then press an arrow key corresponding to the direction of the movement. This could take a long time and so—in a "self-initiated skip" condition—the participant could press both arrow keys simultaneously to skip the current trial and save some time. Conversely, in an "externally-triggered skip" condition, the participant was asked to press both keys upon a certain signal (fixation cross turning red). Authors showed that the self-initiated skips were preceded by an RP-like waveform—which can satisfyingly be explained by the stochastic fluctuation account—but also that compared to the externally-triggered skips, the self-initiated skips were preceded by a notable decrease in RP waveform variability. This demonstrates that the self-initiated actions might not be specifically preceded by an RP (which can be caused by the fluctuations as shown by Schurger et al.,

2012), but they seem to be specifically connected to decrease in EEG voltage variability. Authors conclude that this still can be explained by the stochastic fluctuations, *"with the additional assumption of progressive decrease in the input noise level"* (Khalighinejad et al., 2018, p. 45).

There is one point in Schurger's argument, which I would like to comment on myself—the stochastic fluctuations account does not satisfyingly explain why there is an RP found in Libet et al.'s (1982) pre-set series, when the movement is cued by the clock approaching the target position. If the RP is generated as an artefact of biased averaging, then why would there be an increasing negativity prior to movement which is not internally elicited? One possible explanation—among others, such as the similarity between pre-set RP and CNV—might be that spontaneous fluctuations are not the only cause of movement-preceding negativity and that in the case of pre-set trials, the process is driven by accumulation of both the internal evidence (physiological fluctuations) and the external evidence (clock approaching the target position). I and my colleagues would like to address this issue in the near future by asking the participants to make a movement in a pre-set time, but not necessarily in the first clock revolution—this way, we can analyse whether there is a subliminal negative deflection prior to pre-set time, even though no movement is made.

To conclude these discussions of Libet's experiment—which, of course, could fill a whole book while still leaving out some points—I would like to introduce a recent study, which addressed Libet's concept of conscious veto. This is important, because when discussing whether the RP is truly movement-specific, we also need to ask whether it is possible to stop a movement if the RP already arises (and by extension, to clarify whether RP occurrence leads to voluntary movement inevitably). Libet, Wright, et al. (1983) demonstrated that if the participants are instructed to prepare to make a movement at a pre-set time and then veto the movement at the last instance, the RP is generated but no overt EMG activation is detected. However, this conclusion has several uncertainties; for example, Libet himself (see Libet et al., 1982, p. 330) suggested that RP-like negativity generated in the pre-set tasks might be, at least partially, caused by similarities of these tasks with S1-S2 paradigm which is known to typically elicit CNV (see chapter 2.4.1). Therefore, one could ask whether it is possible to veto a **spontaneous** movement after a readiness potential arises. This question has so far been difficult to answer, because it is almost impossible to detect the RP waveform in a single trial (see the first paragraph in chapter 2.4.1). Nevertheless, a significant progress

in single-trial movement prediction has been made in recent years, eventually leading to an important study by Schultze-Kraft et al. (2016).

Schultze-Kraft and his colleagues (2016) designed an experiment utilizing so called BCI (brain-computer interface), which is a system able to analyse brain activity in real time and use it to communicate with a computer. The experiment was conducted in three phases, while EEG and leg EMG activity were recorded in each of the phases. In the first phase, the participant was asked to step on a button on the floor when a green light was on but stop the movement if the light turned red (which happened at random times). In the second phase, data from the phase 1 were used to predict participant's intention to move—the light turned red as soon as the participant's intention was identified with the aim to make the participant veto the ongoing movement preparation[2]. In the third phase, the participant was informed about the predictive function of the computer and challenged to play a game against the computer, trying to make the movement as unpredictably as possible. The results showed that the participants predominantly could not veto the initiation of the movement if the red light was presented less than 200 ms before the EMG onset. Interestingly, the participants were able to veto the movement if the red light was presented more than 200 ms before the (expected) EMG onset, despite the RP onset was generally found 1000 ms before the EMG onset. This strongly suggests that the movement preparation has a point of no return approximately 200 ms before the movement initiation. In its way, this is compatible with Schurger et al.'s (2012) hypothesis if we assume that the spontaneous fluctuations cross the threshold about 200 ms before the EMG onset. On the other hand, this contradicts Libet's concept of the conscious veto because Libet (1985, p. 537) believed that veto can be exerted only after the W time occurring on average 200 ms before the movement, which would, as Schultze-Kraft et al.'s results suggest, happen too late.

3.5 Sense of agency concept

The Libet's experiment inspired researchers to propose several new concepts. Perhaps the most influential in current literature is the concept of sense of agency. The branch of research pertaining sense of agency usually does not discuss Libet's experiment directly, but its findings offer some important interpretations, relevant especially to the discussion of the

---

[2] It should be noted that the predictive algorithm used was not perfect but managed to predict the movement with an above-chance accuracy.

Study 3 of this thesis (see chapter 6 of this work). Therefore, I would like to briefly introduce sense of agency and several seminal findings pertaining it.

Despite the controversies around neuroscientific research of so-called free will (including the question what we mean by free will), there exists one persistent observation, which is hard to deny—human individuals occasionally feel responsible for their actions. This belief of being an agent is commonly referred to as sense of agency (SoA). For example, Moore, Middleton, Haggard, & Fletcher (2012, p. 1748) define it as follows: "*Sense of agency refers to the sense of initiating and controlling actions in order to influence events in the outside world.*" Sense of agency has a lot in common with Libet's experiment—it also, in its own way, addresses the question of how the subjects experience cognitive processes behind action execution, while relying on their introspective reports.

Sense of agency can be conceptualized as explicit or implicit. The **explicit sense of agency** refers to a high-level consciously accessible impression of being an agent; this can be studied by asking the participants to report the extent, to which they felt to have caused an outcome (for example using a scale, see Chambon, Wenke, Fleming, Prinz, & Haggard, 2013). The **implicit sense of agency**, on the other hand, refers to a low-level feeling of being an agent while the outcomes are *"simply tagged as self-caused or not"* (Moore et al., 2012, p. 1748).

The methods for "measuring" the explicit sense of agency are largely straightforward, such as asking an individual whether an action was intended or not. I say largely, because some methods for measuring the explicit SoA in some studies are not completely intuitive. For example, Moore et al. (2012) consider subjectively reported probability of an outcome to be the manifestation of the explicit SoA. On the other hand, the techniques for measuring the implicit sense of agency are almost always subtler and often notably complex. There are two predominant ways to study the implicit SoA—the intentional binding and the sensory attenuation.

The concept of **intentional binding** was first presented by Haggard, Clark, & Kalogeras (2002) in Nature Neuroscience. Using the Libet's clock, the authors found that if an external event is perceived as an outcome of own voluntary action, then the subjectively reported interval between the action and the consequence is perceived as shorter than if the event follows an involuntary action, in this case, elicited by a transcranial magnetic stimulation (TMS). More specifically, the authors used the Libet's clock to obtain reports of the action

onset (equivalent to the M reports) and reports of the resulting stimulus onset (a form of the S reports); this means that the authors used two individually obtained types of report to estimate the interval length between them. This is not the only way to obtain the interval estimates—given the methodological issues with the Libet's clock, some researchers used an alternative methodology of asking for interval estimates directly (see Moore & Obhi, 2012). Additionally, Haggard et al. (2002) also compared the subjective reports of the stimulus onset in their two conditions (voluntary and involuntary action) with subjectively reported onset of a stimulus occurring after no action whatsoever. They found that compared to this no-action condition, the voluntary action causes the stimulus to be perceived as earlier (i.e. closer to the action, corresponding to the concept of intentional binding), while the involuntary action causes the subjective report of the stimulus onset to be later (i.e. "repulsed" from the action). The introduction of intentional binding raised a wave of interest of other researchers who further developed the methodology and brought a wide variety of evidence supporting the effect (Moore & Obhi, 2012).

Another marker of the implicit sense of agency is the **sensory attenuation**. The concept is based on the fact that a self-caused action leads to attenuation of a resulting stimulus intensity. This was demonstrated by Blakemore, Frith, & Wolpert (1999) who reported attenuated intensity, tickliness and pleasantness for self-caused tactile stimuli, compared to stimuli caused by external factors. Moreover, the authors also showed that the attenuative effect is dependent on the temporal and spatial correspondence between the action and the stimulus by manipulating both the stimulus trajectory and the delay between the action and the stimulus. The sensory attenuation seems to apply as well for auditory modality, as shown by Weiss, Herwig, & Schütz-Bosbach (2011), or visual modality, as recently demonstrated by Vasser, Vuillaume, Cleeremans, & Aru (2019).

Interestingly, Dewey & Knoblich (2014) performed a study of what can be considered a validity of the measures of sense of agency. They reported that the correlation between the intentional binding and the sensory attenuation effects was found to be non-significant, even though both effects were replicated as expected. Furthermore, they did not correlate with the explicit measures of SoA either, suggesting that these methods refer to different cognitive processes.

How does this relate to Libet's experiment? I have already stated that the temporal introspective reports are the key element of Libet's design. However, Libet's main question

is "when" the participant realizes the intention to move, and not "if" the participant feels to be an agent in a given moment. Therefore, it is possible that not all examined movements in Libet's study (and, by extension, in Studies 1, 2 and 3 presented in this thesis) were conducted with the same amount of sense of agency. It is true that Libet asked the participants whether they felt "surprised" by the movement (see Libet, Gleason, et al., 1983, p. 627), and thus employed a form of explicit "measurement" of sense of agency. However, this method of asking was rather rough and little informative. Because a session in the original experiment was considerably long and cognitively demanding, it might be interesting to employ the implicit measures of sense of agency in the future research to explore how many individual movements in such sessions can be considered self-caused.

Empirical part: Libet's experiment validity, replicability and clinical potential

In the previous chapters, I have introduced Libet's experiment and presented several follow-up studies and discussions. In the following chapters, I report our three empirical studies conducted and published in the last three years. It is important to note that these three studies did not have a common general objective; they rather addressed three aspects of Libet's experiment, but their results build up a complex view and hopefully fill some gaps currently debated on the scene of libetian research.

4 Study 1: Validity of measuring the urge to move

In 2017, we published our first study, concerned with introspective judgments of the moment of the first urge to move (Dominik et al., 2017). Majority of the text in this chapter is taken from said publication with the consent of all co-authors (see Appendix 3). The research was motivated by the still open debate presented in the first paragraph of chapter 3.1 of this work, but I will summarize its key points here. Our goal was to test the hypothesis that the W reports are inferred rather than perceived by the participants based on their concept of the M time. Our results strongly support such notion.

4.1 Introduction

The interpretations of Libet's experiment (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, 1985) heavily rely on subjects' introspective reports of their mental states. The most problematic—but also the most important—seems to be the introspective judgment of the moment of the first urge to move.

To obtain these introspective data, Libet and his colleagues (Libet, Gleason, et al., 1983) used the rotating spot method. The researchers placed a subject in front of a modified clock face with a dot on its circumference. The dot was moving clockwise around the clock face, so that it completed one revolution in 2.56 seconds. In the trials relevant to the issue at hand the subject was asked to perform a flexion of fingers or a wrist. After that the subject was asked to state where the dot was when he or she either felt the first urge to move (called W

as in "wanting") or realized the earliest moment of the muscle's actual movement (called M as in "movement").

Validity of the rotating spot method for the M reports seems to be convincingly supported (Pockett & Miller, 2007). Nevertheless, the W reports are much harder to grasp. Unlike M, the W reports do not have a directly observable correlate (which in case of the M reports is the actual movement recorded either by electromyography or simply by a keypress). Therefore, it was argued that the W reports might be artificially induced (Breitmeyer, 1985). Gomes (1998) raised the possibility that the introspective impression of W might not be discernible from the impression of M and that Libet's subjects apparently distinguished between them because they were expected to do so. Interestingly, in his original work Libet had stated that *"Subjects definitively distinguished the experience and time of awareness of wanting to move (W) not only from those of a skin sensation (S) but also from awareness of actually moving (M)"* (Libet, Gleason, et al., 1983, p. 639). On the other hand, Libet admits that he found a suggestive difference in W values based on whether they were measured before or after the M values within respective experimental session (Libet, Gleason, et al., 1983, p. 632).

We postulate that if the W impression is indeed artificially induced, the subjects should **infer** it, rather than introspectively **perceive** it. This means that the W reports might be influenced by factors such as participant's preceding experience or events following the actual movement. Many of these factors seem to be already covered in the literature. For example, Banks & Isham (2009) found an effect of misleading auditory feedback on the W reports, which suggests that the W reports are at least partially dependent on sensory feedback of the movement. Lau et al. (2007) showed a similar effect provided by transcranial magnetic stimulation following the executed movement. They also suggest that the W reports are constructed after the fact, while this process may be influenced by factors following the movement.

Given that Libet himself found an apparent effect of the order of M and W measurements (Libet, Gleason, et al., 1983, p. 632), we expect that the prior experience with M trials might be another of the factors influencing the W reports. Because Libet did not provide a statistical test for this possibility and because he conducted the experiment on only six carefully trained participants, we conducted our own experiment with untrained participants to further explore the nature of the W reports. We propose two hypotheses:

H₁: The M and W reports provided at the beginning of a session do not differ.

This hypothesis reflects the idea than when no preceding M task is experienced, the participants tend to confuse the W reports for the M reports.

H₂: The W reports provided after the M task are earlier than the W reports provided before the M task.

This hypothesis reflects the assumption that the early W reports found in most Libet-style studies are at least partially caused by participants' experience with the M task.

4.2 Materials and methods

4.2.1 Participants

Thirty-five subjects, ranging in age from 19 to 41 ($M = 24.03$, $SD = 5.07$), were recruited in two stages. In the first stage, 19 participants (10 males) were sampled from philosophy students (18 in the sample) and teachers (1 in the sample). Another 16 participants (5 males) were recruited in the second stage via the university Facebook page and posters within university buildings. 13 of these 16 participants were students (in the fields of journalism, philology, nursing, pedagogy, biology, chemistry and history), 2 were librarians and 1 participant reported to be a technical engineer.

All subjects were asked to state how familiar they were with Libet's experiment. Each subject could have stated that he or she (1) had never heard of the experiment, (2) had heard of the experiment, but does not know its procedure or conclusions, (3) knows the procedure or conclusions, but had not read the original study, or (4) had read the original study. The frequencies of each category divided by the recruitment stage can be found in Table 2.

	(1) Had never heard of the experiment	(2) Had heard of the experiment	(3) Knew the procedure or conclusions	(4) Had read the original study
First stage	10	5	4	0
Second stage	12	4	0	0

Table 2: Summary of the frequencies of statements regarding how familiar the subjects in Study 1 were with Libet's experiment. Rows indicate the stage in which respective participants were recruited—first stage contains philosophy students and a philosophy teacher, second stage contains students in various fields, librarians and a technical engineer. Columns indicate the amount of knowledge, which the participants reported.

Responses of the 13 participants reporting any prior knowledge of the experiment (i.e. participants in categories 2, 3 and 4 in Table 2) were excluded from the analyses, as we aimed to analyse data only from naïve individuals. Another 2 of the 22 naïve participants were excluded from the analyses due to technical issues and potential misunderstanding of the instructions. All data reported further in this chapter therefore describe the sample of the 20 remaining participants.

Nineteen of the uninformed subjects stated that their dominant hand is right, one participant reported the left hand to be dominant. The participants were also asked how energetic they were feeling: no participant reported feeling very tired, 3 participants reported feeling tired, 11 participants normal, 6 energetic and no participant reported feeling very energetic. We did not exclude any subjects based on how energetic they were feeling.

The experiment was conducted on four separate days, while each participant attended only one approximately 60 minutes long session. Each participant had equal probability of being assigned to either of the experimental conditions regardless of the stage he or she was recruited in. All participants gave informed consent in a brief preliminary questionnaire and as a reward, they were offered a subsequent tour through a virtual reality lab with the opportunity to experience the virtual reality with a head-mounted display and motion sensors.

4.2.2 Technical equipment

We used 19 Internet connected office computers. Our own online computer program—specifically designed to perform all types of tasks from Libet's original experiment (see chapter 5.2.2)—was ran in an Internet browser environment on each computer. This online solution allowed us to perform the experiment on many computers simultaneously. For purposes of this study, only **the M and W reports in the absolute mode of recall** were collected (see chapter 2.3). The clock outline was white on black background (see Figure 2b in chapter 2.2.3). There was a white spot in the middle of the clock where the subjects were supposed to fix their gaze. Another white spot was moving clockwise on the inner side of the clock's circumference. The moving spot completed one revolution in 2.56 seconds, which is identical speed to Libet's original study. Subjects were asked to press a key whenever they wanted to. After the keypress, they were supposed to report the M or W time with a mouse-click on a certain position on the clock's circumference.

4.2.3 Experimental procedure and design

The experiment was conducted in a computer lab at the Department of Psychology at Palacký University Olomouc, the Czech Republic. After the arrival, each subject drew a card from an opaque sack containing a unique participant ID and a codename of one of two groups into which each individual was assigned. This method was used to assure random assignment of the participants to the experimental conditions. The participants were then seated in front of a computer screen and asked to fill in a brief online questionnaire concerning their current self-perceived energy state and some personal information, such as a field of study or work and how deeply familiar they are with Libet's experiment. Once all the subjects finished, they were asked to proceed to the main task.

After making some personalization settings (such as filling in subject's ID, handedness and experimental group), the instruction was displayed. All participants were required to place one of their hands on a key—right hand on the numerical Enter key if right-handed or left hand on the left Ctrl key if left-handed—and fix their gaze on the spot in the middle of the clock. After the spot on the clock's circumference started moving, the subjects were asked to let it finish at least one whole revolution. Subsequently, while still fixing their gaze on the central spot, the subjects were supposed to press a key in whichever moment they wanted to. It was emphasized that the movement should be brief and spontaneous, made in the moment of the first urge to do so. After the keypress, the spot on the circumference continued moving for a short period of time, after which it disappeared. The subject was then asked to mouse-click on the position where the dot was in the moment in which he or she realized either (1) the first **urge** to move (time W) or (2) the start of the actual **movement** of the finger (time M). After the response, the subject clicked on "OK" button and repeated the whole process. The participants were advised not to rush through the experiment and use some time to blink or close their eyes to relax before they clicked "OK". Specific instruction about blinking during the experiment was given—subjects were asked not to blink during the clock running; if the urge to blink became uncomfortable, the subjects were allowed to blink, but then wait at least one full revolution before proceeding. To review the whole instruction, see Appendix 4.

Each subject performed 40 trials reporting one type of values (M or W) and then another 40 trials reporting the other type of values. After finishing the task, they were directed to a short YouTube video explaining Libet's experiment (BBC Radio 4, 2014). Subjects had been

asked in the invitational email to bring their own headphones, so that they could watch the video without disturbing others. Once all the subjects finished, they were debriefed, all the questions were answered and the tour through the virtual reality lab was provided. The subjects were also given the opportunity to send us their IDs (naturally while giving up the otherwise guaranteed anonymity), so that we could send back their personal results in the experiment.

4.2.4 Variables

We aimed to examine the idea that the W values are artificial and in fact just a result of subjects' inferences based at least partially on the time M. To provide evidence, we created a two-group between-subject design with following independent variables or factors:

(1) **Condition.** This factor distinguishes simply between M and W values. As Libet showed in his results, there already seems to be a substantial difference between these two reported times (Libet, Gleason, et al., 1983). However, Libet's results were obtained after a thorough training, so all his subjects had already had an experience with both M and W measurements. Therefore, we are especially interested in the difference between these two values when they are measured as first in the session, i.e. with no effect of preceding session of any kind (H_1).

(2) **Order.** In our design, both groups of participants perform the same tasks, but the subjects in one group (group A) perform the W task before the M task and the subjects in the second group (group B) perform M before W. By examining the interaction between *Condition* and *Order*, we may observe any effect, which the preceding M task might have on subsequently reported W values (H_2).

(3) **Subject.** Because each participant performed 80 trials (40 in W condition and 40 in M condition), we introduced a random variable called Subject, which allows us to account for the fact that the responses may differ simply because there are multiple different participants.

The dependent variable is represented by the reported times (in ms) measured in each trial. Negative time means that the event in question happened before the keypress.

4.2.5 Data analysis

The analysis of the effect of the factors and interaction between *Condition* and *Order* made use of the linear mixed-effect model. The rationale behind choosing this method was

the hierarchical arrangement of factors. Levels of the variable Subject are nested within the *Order* factor (each subject could have been assigned to only one level of the *Order*). It makes the regressors fully linearly dependent and therefore their effects are impossible to be estimated with the ordinary method of least squares. A schema of our model is depicted in Figure 5.
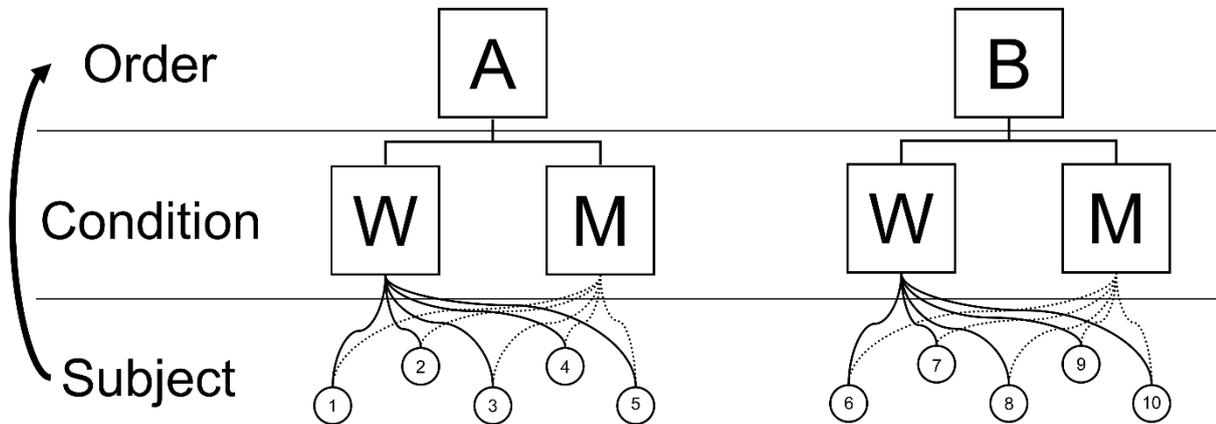
Figure 5: Schema of overall linear model factor structure. Example subjects 1 to 5 are members of group A, example subjects 6 to 10 are members of group B. Each subject performs 40 trials of both M and W series, as well as each order (A or B) contains both M and W series. A bold arrow is used to signify that the factor Subject is nested inside the factor *Order*.

The model consisted of the fixed factors *Order*, *Condition* and their interaction term. The random factor Subject enabled individual respondents to differ in intercept within the M and W condition separately and allowed magnitudes of the W effect to be individual for each participant[3]. We hypothesize that the *Order* factor does not influence the reports made in *Condition* M, but only those made in *Condition* W. Therefore, we include the interaction term, which covers only the responses made under these two circumstances: a) the *Condition* is W and b) the subject has already encountered the M condition (represented as *Order* B). We expect the interaction term to have significantly negative effect. More clearly, we hypothesize that those subjects who are aware of the existence of M condition locate their W responses in earlier times than others.

The factors *Order* and *Condition* divide the measured values into four clusters. The following notations will be used:

---

[3] As for the nuts and bolts of the analysis, packages lme4, lmerTest and BayesFactor of the R statistical software were used. The model was specified with a formula Time ~ Order * Condition + (1+Condition|Subject). Shatterthwaite approximations to degrees of freedom were employed in significance tests of dependent variables.

(1) **W$_A$**. Values measured in W condition in group A, therefore measured first in the session.

(2) **M$_A$**. Values measured in M condition in group A, therefore measured second in the session.

(3) **M$_B$**. Values measured in M condition in group B, therefore measured first in the session.

(4) **W$_B$**. Values measured in W condition in group B, therefore measured second in the session.

Besides the exclusion of the 15 participants as stated in chapter 4.2.1, the data were checked for outliers. The means of reporting time using the clock has its limitation in the case that a participant reports extremely early time which appears as a late time of the previous revolution of the clock. The same distortion can be observed in extremely late times appearing as early ones. To eliminate this distortion, we employed a script to check each subject in both conditions separately and position the revolution breakpoint to the most suitable time (i.e. to the largest gap between times). Subsequently the outlying observations identified with the Tukey's 1.5×IQR rule were removed from each set of 40 trials. This procedure led to omitting 67 out of 1600 (4.2%) values removing from 0 to 5 observations from each set. It should be added that skipping the data cleaning procedures only has little effect on the results of the statistical tests and the estimated weights. However, without the outlier removal the model residuals do not meet the assumption of the normal distribution.

Our hypothesis H$_1$ expects that there is a null effect. Unfortunately, statistical tools for testing the null hypothesis are scarce. The failure to reject the null hypothesis cannot be interpreted as evidence for non-existence of the effect rather than absence of evidence for its existence. That is why we have employed a statistic called the Bayes factor (BF) besides each significance test. The Bayes factor is a statistical index that quantifies the strength of evidence for a hypothesis, when compared to another one. It is calculated as a ratio of the likelihood of two competing models: the constrained model in denominator representing the null hypothesis and the relaxed model in numerator standing for the alternative. The Bayes factor states how many times more likely it is to get the observed data if the alternative is true than if the null hypothesis is true. Therefore, the value of the index higher than 1 favours the alternative and conversely value lower than 1 favours the null hypothesis (for details see Ly, Verhagen, & Wagenmakers, 2016). Generally, the index values between 1/3 and 3 are

considered to provide a support that is "not worth more than a bare mention" (Robert, Chopin, & Rousseau, 2009, p. 18) for any hypothesis.

4.3 Results

The statistical model showed no significant main effect of the factor Order, $t(19.99) = 0.634$, $p = 0.533$, $BF = 0.82$. This is in accord with our assumption that there should be no effect of the order on both M and W values, even though the Bayes factor fails to provide evidence for its non-existence. Additionally, no significant main effect of the factor Condition was found, but the Bayes factor inconclusively favours its existence, $t(20.02) = 0.665$, $p = 0.514$, $BF = 2.25$. This result, however, is still interesting due to the fact that the difference between M and W values is expected to be quite substantial (see Libet, Gleason, et al., 1983). As expected, we found a significant effect of the interaction Order*Condition, $t(20.02) = -3.488$, $p = 0.002$, $BF = 9.24$. This means that the factor Order influences M and W values differently. The results are summarized in Table 2.

Factor	Estimate	t value (df)	p value	Bayes factor
Intercept	-52			
Order (B)	44	0.634 (19.99)	0.533	0.82
Condition (W)	41	0.665 (20.02)	0.514	2.25
Order*Condition (B*W)	-319	-3.488 (20.02)	0.002**	9.24

Table 3: Summary of the significance of individual factors. The Intercept represents the mean value of the reported M times in group A. The Order quantifies the difference between the mean time reported in order A and B (positive value indicates later times when order is B). The Condition represents the difference between mean M and W reports (positive value indicates later times in W condition). The interaction term Order*Condition quantifies additional effect present in times reported under condition W and order B. Double asterisk indicates significance at $\alpha = 0.01$.

Within the H_1, we are interested in the difference between the M and W values when they are measured as first in the session. In our case, these are the values W_A and M_B. No significant difference between the two reported times was found, while the Bayes factor approaches the 1/3 threshold, $t(31.63) = 0.042$, $p = 0.967$, $BF = 0.41$. This further raises our suspicion that there might be no significant difference between M and W values when no previous training is performed, even though our analysis does not provide conclusive evidence. For the purposes of testing the H_2, we examined the difference between the W_A and W_B values to analyse the possible effect of the experience with previous M sessions on the W_B values. We found a significant difference, while the W_B values were significantly

earlier than the W_A values, $t(20.00) = -2.75$, $p = 0.012$, BF = 11.28, hence clearly supporting our hypothesis. For an overview of the clusters, see Figure 6.
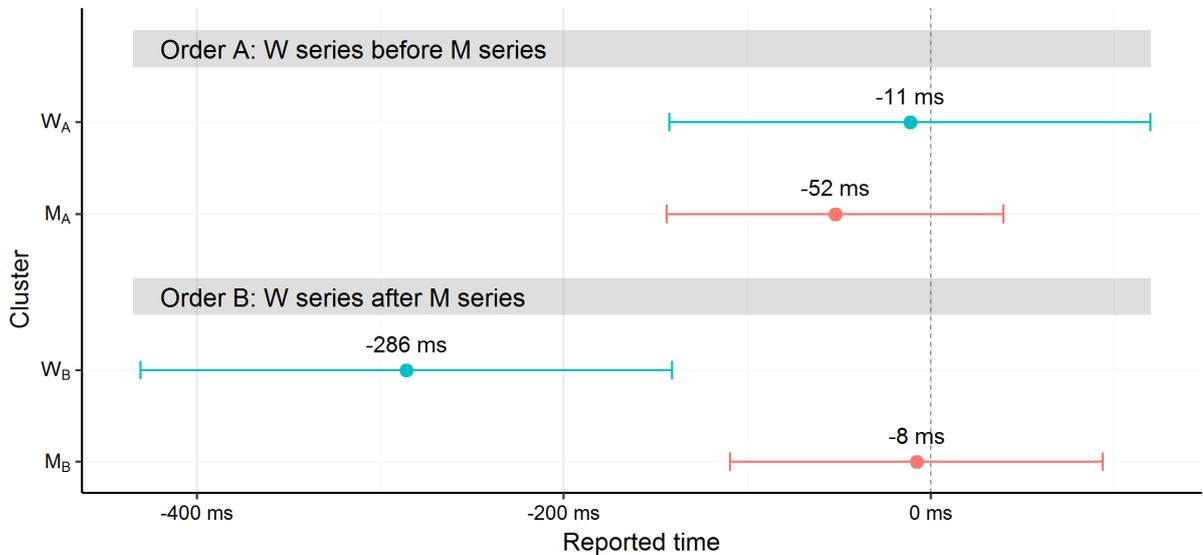
Figure 6: Reported times in Study 1 by clusters. The dots with a value above them signify the mean values for each cluster of reports. The error bars show 95% confidence intervals for expected values. W_B cluster differs significantly from W_A, $t(20.00) = -2.75$, $p = 0.012$, BF = 11.28, M_A, $t(29.44) = -2.67$, $p = 0.012$, BF = 4.78 and M_B, $t(20.01) = -4.10$, $p < 0.001$, BF = 38.46. No other differences are significant: W_A M_A, $t(20.02) = 0.665$, $p = 0.510$, BF = 0.37, W_A M_B, $t(31.63) = 0.042$, $p = 0.967$, BF = 0.41, M_A M_B, $t(19.99) = 0.634$, $p = 0.533$, BF = 0.48. No correction for multiple significance tests / confidence intervals was employed.

To summarize our findings in the context of the proposed hypotheses:

H_1: The M and W reports provided at the beginning of a session do not differ.

The hypothesis was not conclusively accepted, but a suggestive trend in its favour was found, BF = 0.41.

H_2: The W reports provided after the M task are earlier than the W reports provided before the M task.

Hypothesis was accepted, $p = 0.012$, BF = 11.28.

4.4 Discussion

Libet and his team (Libet, Gleason, et al., 1983) claimed that their subjects had no difficulties distinguishing between M and W values while reporting them. Nevertheless, a possible explanation arises that this might be caused by the previous training of the participants. If we were to assume that there exists no such impression as "the first urge" or

"wish to move", we would expect that the apparent ability to distinguish W from M might be based simply on the participants' effort to comply with the experimenter's expectations.

Our results suggest that the difference between M and W values when measured at the beginning of the respective sessions might be non-existent, but this evidence remains inconclusive and needs another testing on a larger sample. Nevertheless, it seems that when the subjects are confronted with the task for the first time, they tend to report similar time on the clock regardless of whether they are asked to report the impression of the first urge to move (W) or the impression of the movement itself (M). This is in accord with Gomes's (1998) supposition that the impressions observed in Libet's experiment might be a manifestation of the same mental event.

It is important to note that Libet replied to Gomes's objections. In the reply in the section addressing this particular issue, he repeats that his statements were based *"(a) on the subjects' own expressions that they had no difficulty in making this distinction and (b) on the statistical reliability of the differently reported times for W vs. M series. The differences between values reported for W and M were substantial and statistically very significant"* (Libet, 2000, p. 9) These arguments seem to be weakened by the fact that all these subjects were already deeply familiar with the task in both M and W condition. We suggest that there seems to be a significant effect of the mere experience with M measurements subsequently *producing* the difference between M and W values. If that is true, we might assume that what Libet observed was not a difference between two introspective phenomena, but a difference between one introspective phenomenon and one artificially induced impression.

Our analysis also showed that there is a significant difference between W values measured first in the session and W values measured after the M values. Specifically, the W values measured after M (in group B) were much more negative (on average -286 ms in group B while the average time W in group A was -11 ms). This suggests that the previous experience with M measurements—or maybe any other kind of measurements, given that we did not test this possibility—decreases the subsequently measured W values. This may be explained by an idea which, for purposes of this discussion, we call "anchoring". When the subject performs the W task first, he or she is not aware that another task will follow requiring to report the time of the impression of the actual movement. In accord with our assumption that there might be no difference between the notion of the urge to move and the movement itself, the subject simply reports the time of the movement as W. However, subjects in the second

group perform the M task first, and after that they proceed to the W measurements. Because they are already familiar with the time, which they assume to be the time of the actual movement (i.e. they are "anchored"), they tend to report significantly earlier W times, because the impression of the urge to move should logically occur before the impression of the actual movement. Pockett and Purdy (2011, p. 38) illustrate this issue by asking whether the subjects *"infer after the event that, because the experimenter asked about their urge, etc., they must have had one—and it must have occurred a bit before the movement—which puts it probably about… there…?"* This explanation raises the possibility that the W values are not *perceived*, but rather *inferred* from the subjects' conception of the time of the actual movement.

It seems that studies aimed on this particular issue are scarce. Somewhat similar to our design seems to be a recent study of Caspar and Cleeremans (2015). Their experiment was aimed primarily to examine how some psychiatric issues or personality traits might relate to the reported W times[4] and measured RPs and LRPs. Their results regarding the W times showed significant main effect of an equivalent to our factor Condition, but no significant effect of the Order or the interaction Order*Condition. These results seem to contradict our findings. In order to eliminate the risk that the different results are caused by different statistical approaches, we requested the data from the authors and analysed them using the identical procedure as described in chapter 4.2.5, including the outlier removal rule. The results matched the original Caspar and Cleeremans's (2015) outcome—significant effect was found for the Condition factor only, even though the Bayes factor does not provide a conclusive support for the alternative hypothesis, $t(92.75) = -8.07$, $p < 0.001$, $BF = 1.50$. The effect of the factor Order was virtually zero, $t(92.61) = -0.17$, $p = 0.863$, $BF = 0.12$, as well as the effect of the interaction Order*Condition, $t(92.79) = -0.21$, $p = 0.836$, $BF = 0.14$. The intercept equals -59 ms. This shows that we can acquire similar results as the original authors, even while using the mixed-linear model. However, because the different statistical procedure cannot account for the difference in results between our and Caspar and Cleeremans's study, we need to provide an explanation for it. In personal communication, the authors stated that their participants were naïve before entering the experimental session. Nevertheless, because the subjects performed 8 training trials at the beginning of the

---

[4] These findings constitute the basis for Study 3 of this thesis.

experiment (4 in W and 4 in M condition), their naivety might have decreased by the time they started performing the regular task. This fact might account for why Caspar and Cleeremans's results show clear difference between the M and W times even while measured at the beginning of the respective sessions.

Our interpretation is in accord with the view of Banks and Isham (2009) who provided experimental evidence suggesting that the reported W times are inferred rather than perceived. Their experiments showed that the W reports might be significantly influenced by a misleading auditory or visual feedback while the feedback is applied after the movement. The authors argue that *"(...) a large component, possibly the entire estimate, of W is retrospectively inferred from the response (...)"* (Banks & Isham, 2009, p. 19), which corresponds with our interpretation. It is possible that the feedback provided by Banks and Isham may provide the "anchoring" similar to the effect caused by the preceding M series in our experiment, even though in the opposite direction. Similarly, our results correspond with the study of Lau et al. (2007), who also demonstrated that the W reports are prone to be biased by post-movement factors.

Our conclusions also compliment the results of Pockett and Purdy (2011) who suggest that the urge or wanting to move might not be experienced, but rather inferred. Like in our case, Pockett and Purdy examined the M and W reports of untrained participants and noticed that (perhaps due to high inaccuracy of the W reports) they did not find significant difference between the M and W reports. More remarkably, one of the subjects reportedly claimed that he was unable to tell the difference between wanting to move and actually moving.

The findings presented here are also potentially applicable to clarify some unexplained results of other studies. For example, Trevena & Miller (2002) reported remarkably late W reports compared to what they expected. The authors concluded that the participants might have simply been wrong about the reported W time. Nevertheless, given that it seems that no M reports were collected within their experiment, our hypothesis provides a plausible explanation: Trevena's and Miller's participants might have reported late W times, because they **did not have the prior experience with reporting the M times**. Similarly, Miller et al. (2011) found that their participants reported the W times on average 30 ms after a keypress and interpreted it as a sign of the participant being able to *"generate movements that were extremely spontaneous with respect to both conscious awareness and brain processes of movement preparation—just as they were instructed to do"* (Miller et al., 2011,

p. 108). According to our hypothesis, this might be simply explained by the fact that the participants did not experience the M task. Admittedly, however, I point out in chapters 5.3.4 and 5.4.4 of this work that these remarkably late W reports appeared in our results of Study 2 as well, even though our participants in this study did have prior experience with the M task. Therefore, the explanation offered here is not definitive and more research is needed.

It was also recently noted in Lush & Dienes (2019) that our hypothesis can be supported by the results of Lush, Naish, & Dienes (2016), who investigated the subjective timing of intention to move in hypnotizable and meditating respondents. The research consisted of three studies, some of which included the M reports in addition to the W tasks and some did not, while the results correspond with ours.

A discussion of the overall importance of the findings might come into question. Even if the W reports are indeed inferred, what does it mean for the general results of Libet's original experiment and the experiments following its paradigm? Our results suggest that it is incorrect to compare the W values to the RP or any other objective data. In the original experiment of Libet, Gleason, et al. (1983), the W time was supposed to indicate the first urge or wish to move available to the subject's awareness. Therefore, it made sense to compare W to the RP onsets, as both might have been considered the potential cause of the movement. However, because our data suggest that the subjects might have inferred the W reports without actually perceiving the impression, the whole existence of such experience as "the first wish to move" is in question.

4.5 Limitations and Recommendations

Our experiment is based on the data obtained from naïve participants with no training whatsoever. This particularity of our research makes the results difficult to compare to other studies' outcomes. As suggested by the comparison of our results to those of Caspar and Cleeremans (2015), it is possible that even a very brief training session containing only four trials in each condition might create a difference between M and W values. However, there seems to be no other way to test our hypotheses than by testing untrained subjects.

The reliability of our results is weakened due to wide confidence intervals, especially in the case of the W reports. Also, a support for the hypotheses that some differences are non-existent is inconclusive. A similar experimental design with larger sample size may overcome these issues and provide more reliable results.

From the technical point of view, the accuracy of our apparatus might be in question due to a risk of lag given that the task was performed online. Nevertheless, our program performed all the operations on the computer locally and only complete results were sent into central database.

Some untreated confounding variables were also present in the procedure, such as both types of handedness, variable distance of the subjects' eyes from the display or possible differences in the technique of the keypress. However, these individual differences probably only have a minor effect unrelated to our findings, as the linear model accounted for the possible interindividual variability.

We have taken into account the possibly confounding variable whether the subjects are familiar with Libet's experiment. It led to the exclusion of 13 subjects from the analysis. However, this presumable confounder would not affect the results significantly even if it had not been considered. The results of the analysis of all subjects together are almost identical to the results presented in this study. The new factor indicating whether the subject is familiar with the experiment is not significant, $t(32.98) = 0.658$, $p = 0.515$, $BF = 0.40$, and neither is its interaction with the factor Condition, $t(33.00) = -0.612$, $p = 0.545$, $BF = 0.38$.

We did not test whether the W values are influenced by preceding M measurements specifically or by any other task using the rotating spot method. Therefore, it might prove useful to include another experimental group performing dummy task instead of the preceding M measurements. It might also be interesting to repeat the measurements about one week later using the same order, which might reveal whether the W values stay significantly lower than M values or whether the difference between M and W values increases over time. Moreover, a repeated measurement after some time may be informative for Libet's original experiment, as his participants were tested on a roughly weekly basis.

4.6 Conclusions

The experiment suggests that there might be no difference between the reports of the first intention to move (W) and the impression of the movement itself (M) when both are measured at the very beginning of the experimental session, which may imply that the difference appears only after several training trials. Additionally, the W reports obtained at the beginning of the session differ significantly from the W reports obtained after the M measurements. This may imply that the W reports are not directly introspectively perceived but rather inferred at least partially from the previous experience with the M measurements.

5 Study 2: A complex replication of Libet's experiment

In the previous chapter, I have discussed our study specific to introspective component of Libet's experiment. Nevertheless, as was shown in the chapter 3, Libet's experiment is being debated with regards to almost all its aspects. This stimulated us to conduct a large-scale replication study (to our knowledge first of its kind), verify its results and pinpoint some problematic features relevant to modern-day Libet-style experiments. The text of this chapter is mostly taken from our 2018 publication (Dominik et al., 2018a) with the consent of all co-authors (see Appendix 3).

5.1 Introduction

Given its technical complexity and interpretational impact, one would expect that there would be a large initiative to replicate Libet's experiment soon after its publication. Nevertheless, the empirical studies replicating the original experiment in some way or another seem to adopt a different approach. Vast majority of these studies substantially modified or simplified the experimental methodology to either provide support for individual counterarguments against Libet's conclusions (e.g. Keller & Heckhausen, 1990; Trevena & Miller, 2002; Schurger et al., 2012; Verbaarschot et al., 2015) or expand the study using more advanced technology and procedures (for example using an fMRI machine, see Lau et al., 2004; Lau et al., 2006; Soon et al., 2008).

The more recent the studies are, the more they seem to be focused on specific aspects of the original Libet's experiment. For example, a study made by Danquah et al. (2008) aimed at the tactile stimulation which Libet used in his S series (see chapter 2.3.3) and showed that the tactile stimuli may be consciously registered with different latency than visual or auditory stimuli.

Trevena & Miller (2002) presented two experiments focused mainly on analysing lateralized readiness potentials (LRPs) instead of the classical readiness potentials (RPs) used in Libet's case. These authors modified the original design and ultimately had to admit that the participants' introspective reports of the timing of conscious decision to move differed substantially from what was found by other researchers, including Libet, and that some of these W times were reported after the movement was initiated. I have suggested in the previous chapter that this might have been caused by the omission of the M tasks in their experiment.

Schurger et al. (2012) introduced a modification, which they called "Libetus interruptus" design, and showed that the early readiness potentials' onsets observed by Libet and his team may be caused by spontaneous fluctuations in neural activity occasionally building up to a movement execution.

Other studies focused on biases in reporting the introspective impressions in Libet-style experiments (e.g. Pockett & Miller, 2007; Banks & Isham, 2009; Pockett & Purdy, 2011). These—together with our Study 1 reported in chapter 4—showed that certain types of introspective reports (especially the reports of the urge to move) are highly susceptible to being distorted due to changes in the experimental situation.

These and other similar studies are extremely informative for isolated aspects of Libet's experiment. Nevertheless, their experimental designs are usually notably reduced and do not reflect the complexity of Libet's original study, which suggests that it might prove useful to conduct a complex replication. That means a study which does not aim to challenge Libet's results or interpretations, but instead attempts to conduct the original experiment following Libet's methodological directions as closely as reasonable.

That is the aim of the present study. Our procedure consisted of four general steps as follows: (1) we familiarized ourselves with Libet's original methodological papers (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, 1985), (2) we devised our own technical plan of the experiment following the original directions, (3) we enhanced the design slightly to overcome some of its original methodological limitations and to adjust it to equipment available to us, and finally (4) conducted the experiment including the data analysis. To maximize the benefits of our study, we published the research data along it to allow other researchers to revise an authentic data sample (see Dominik et al., 2018b). Due to its size, the data sample is not part of this thesis, but can be found in mentioned publication.

One could argue that our study cannot be called a replication in the strictest sense, since we decided to make some changes to the design. While this might certainly be true, in case of such a complex experiment it is often difficult to balance replicative accuracy and methodological generalizability, ultimately forcing the researchers to make choices between keeping the design intact, but less valid, and improving it so that it is more valid, but less accurately reproduced. In short, while we were aware that modifying the design may lead to changes in the results, we found some modifications necessary, either for technical or for methodological reasons.

5.2 Materials and methods

5.2.1 Participants

Our research sample consisted of eight participants. This increase in sample size is not large compared to Libet's six participants, but our number of eight has a rational reason: it allows a complete rotation of three experimental conditions, which could not be satisfyingly rotated if we examined six participants only (see chapter 5.2.5). Our participants were recruited from undergraduate psychology students during November and December 2015.

The recruitment consisted of three steps. In the first step, the potential respondents reacted to an offer distributed via university email and webpage by filling out a questionnaire containing items such as name, contact information, gender, handedness, age, approximate hair length (relevant to the quality of prospective EEG signal measurement), near-sightedness and so on. In the second step, we chose 13 participants with convenient answers in the questionnaire and invited them to a group briefing session, in which we introduced them to the research idea, time requirements, ethical regards and a financial reward. To the question whether the students can search for more information on the original experiment we replied that we prefer to introduce the participants gradually ourselves. However, all participants were clearly instructed, informed and trained before any measurements were taken, so the participants' awareness of the details of the experiment should not have a major effect on the results. In the final step of the recruitment, we randomly chose 8 participants to obtain data from, with respect to our methodological requirements (4 females, 4 males, right-handed, full-time students only).

These 8 students were assigned into two groups of 4. The first group was studied from January to March 2016; the second group was studied from March to May 2016. The interval between these two study periods was two weeks.

5.2.2 Introspective data measurements

To obtain the introspective data, we used an office computer and a laptop, both running a custom web-based program with Libet's clock. The program is a traditional rotating spot method designed to be used in various Libet-style experiments (such as our experiment in Study 1), but its properties strictly follow Libet's original recommendations (Libet et al., 1982, p. 324; Libet, Gleason, et al., 1983, p. 625). Since Libet used a CRO display, which had a bright green outline of a clock face displayed on dark background, our clock face was

displayed as a white outline on black background to maximize the contrast. The outline was labelled by 12 large marks, 11 of which were numbered by increments of 5. In the middle of each pair of adjacent numbered marks, a smaller mark was displayed (see Figure 2b in chapter 2.2.3). A dot on the clock's circumference moved clockwise and completed one revolution in 2560 ms. The clock diameter and its distance from participant's eyes were adjusted, so that the viewing angle never exceeded 1.8° (in fact, we aimed to keep the viewing angle at exactly 1.8°).

The computer program first loaded an introduction page with the following option settings: session ID, participant's name, gender, session scenario, notes, clock diameter and some other options, which were, however, set always the same (such as right handedness or clock speed). Once the introduction page was filled out, an instruction was displayed (the text varied based on the session scenario and participant's gender, see Appendix 5). When the participant finished reading the instructions, he or she pressed and held the mouse button, initiating the following task and marking the beginning of the task in the recordings (see chapter 5.2.3).

The tasks varied based on the session scenario (see chapter 5.2.5). Generally, the participants were asked to either make a movement (click the left mouse button) or wait for a skin stimulus to be delivered while watching the centre of the running clock. After either of these specific events occurred, the moving dot on the clock face continued moving for a continuation interval (see Libet, Gleason, et al., 1983, p. 626) of random length ranging from 500 to 800 ms after the event, and then disappeared. Thereafter, participants were asked to report the location of the moving dot when they registered a specific introspective impression. This report could be made by two different modes of recall, as proposed by Libet, Gleason, et al. (1983, pp. 626–627). In our design, the absolute mode of recall (A) meant that the participants simply clicked on a specific spot on the clock face where the dot was when their impression first occurred. The order mode of recall (O) used a more complex way of reporting described in chapter 2.3 of this work. To summarize, after the continuation interval, the moving dot jumped to a specific stop time. Its position was chosen from 41 possible values from the stopping range, which was spread across an interval from 400 ms before the event to 200 ms after the event (in our case, no value was omitted; the number of trials was increased to 41 instead). The participants were then asked to compare the timing of the subjective impression with the current dot's location. In this arrangement, three

eventualities might occur: (1) the participants' impression comes before the stop time ("awareness first", scored as 1 point), (2) after the stop time ("clock first", scored as 0 points) or (3) exactly at the stop time ("together", scored as 1/2 point). Participants stated their answers by clicking on one of three boxes labelled by these three eventualities (the scores were not displayed). The final average time of the awareness was then calculated using the following formula stated also in chapter 2.3 of this work:

(upper, positive end of "stopping range") – (time interval between "stop times") × (number of points – ½)

5.2.3 Recordings and skin stimulation

Besides measuring the introspective reports, Libet also recorded physiological data, namely EEG and EMG. EOG had also been employed in the first few sessions to account for unwanted eye movements, but was later discontinued, because the EOG potentials rarely occurred, as it seemed that participants could satisfyingly fixate a mark in the centre of the clock (Libet et al., 1982, p. 323). Therefore, we decided to record EEG and EMG only, as the EOG would presumably pose unnecessary additional discomfort to our participants.

For both the EEG and EMG recordings, we used the BIOPAC MP150 unit with EEG100C and EMG100C amplifiers. The EEG recordings were taken to obtain data on the readiness potential onsets. We recorded the EEG from six standardized 10-20 locations using 0.1 Hz high-pass and 35 Hz low-pass filters. The Fp_1, Fp_2, C_z and P_3 electrodes were recorded in accordance with Libet's directions (Libet et al., 1982). However, Libet also recorded additional non-standard electrodes C_c and C_i (Libet et al., 1982, p. 323), which are not included in the 10-20 system. We decided to record C_3 and C_4 electrodes instead, because the EEG cap available to us does not allow recording of unstandardized locations outside of 10-20 system. The electrodes were embedded in the BIOPAC CAP100C electroencephalography cap filled with an electroconductive gel. Reference electrodes were placed on both left and right ear lobes. For the overview of the electrode sites, see Figure 7.

Because the movement, which our participants were asked to conduct, was a mouse click, we recorded the EMG from the musculus extensor indicis using two EL503 Ag/AgCl electrodes applied to the skin abraded with ELPREP gel. The ground electrode was located on the upper part of the musculus brachioradialis. We applied 10 Hz high-pass and 500 Hz low-pass filters.
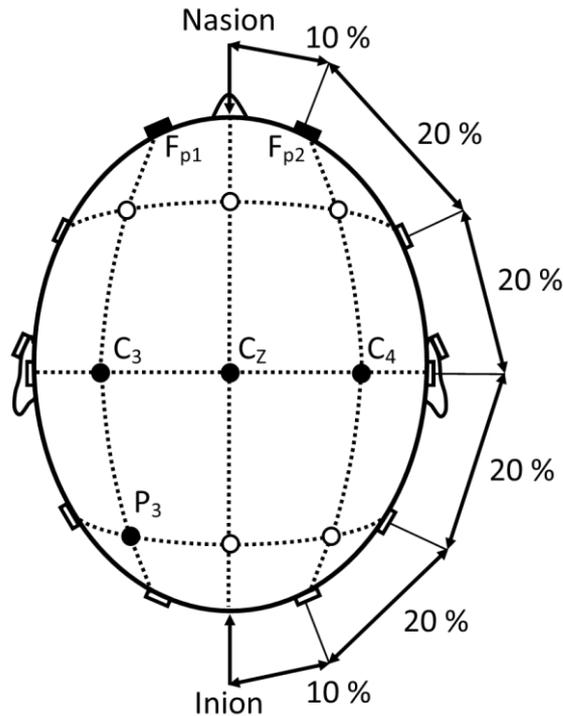
Figure 7: EEG electrodes sites on participants scalp in our replication. Electrodes C_c and C_i were replaced by C_3 and C_4.

Both the EEG and EMG data were obtained, stored and pre-processed in the AcqKnowledge software v4.4. The AcqKnowledge also recorded the mouse clicks as a digital input reaching 1 when the mouse button was being pressed and returning to 0 when it was released. This was achieved by connecting the computer running Libet's clock to the BIOPAC STP100C module via a CBL110C cable plugged into the LPT port of the computer. To convert a click to an output signal in the LPT port, we used a simple custom script. This arrangement was utilized to accurately align the timeline in Libet's clock with the timeline of the physiological recordings, as each click also created a timestamp in Libet's clock.

Another important part of Libet's original experiment was the skin stimulation (S) series (Libet, Gleason, et al., 1983, pp. 627–628) employing an electrical stimulator placed on the back of participant's hand (Libet et al., 1982, pp. 323–324). We used a non-electrical tactile stimulator TSD190 provided by BIOPAC connected to the STM100C module set on CH15 source setting. The stimulator was placed on the anterior side of participant's left wrist. It was controlled by an analog signal sent from a laptop via a 3.5mm audio cable to channel 16 on the UIM100C module. For safety reasons, we used a battery-powered laptop—connected to the same display which was used in all other types of tasks—instead of an office PC. The stimulus intensity was adjusted at the beginning of each session by reducing it to a level at

which participant could no longer feel the stimulus and subsequently increasing the intensity slightly by about 15% of a turn of the cylindrical slider on the STM100C. The stimulus was thus *"sufficiently weak to make recognition somewhat difficult, but sufficiently above threshold to eliminate equivocation about stimulus delivery"* (Libet et al., 1982, p. 324). The temporal alignment of Libet's clock with the AcqKnowledge timeline was in this case achieved using the skin stimuli themselves, because the STP100C module—otherwise registering the mouse clicks—had to be disconnected from the rest of the system for technical reasons.

5.2.4 Types of the experimental tasks

The explanation of the various tasks present in Libet's experiment can be found in chapter 2.3 of this thesis. In the present chapter, I will discuss our methods for these tasks in our replication.

The **M series** in our replication consisted of 40 trials in mode A or 41 trials in mode O. During each trial, the participant was seated in a medical armchair. A wooden board was placed on the chair's armrests and a computer mouse was put on top of it. The participant laid his or her right hand on the mouse and fixated the mark in the centre of Libet's clock. The participant was asked in the textual instruction to click the left mouse button whenever he or she wanted (see Appendices 5.3 and 5.4). The click was followed by the continuation interval after which a prompt was displayed, asking the participant to report the time of subjective recall of the beginning of the movement (M). The prompt was different for A and O modes of recall (see chapter 5.2.2). Participants were advised to look at a wall after making the report, if they felt the need to relieve their eyes; and then click an OK button to continue. Participants were also instructed not to blink while the clock was running, unless the urge to blink became uncomfortable; in such a case, participants could blink, but then had to wait at least one whole clock revolution before making another click.

The **W series** was almost identical to the M series, with a slight difference in the introspective event reported after the click. In the M series, participants reported when they first realized that they were moving their finger. In the W series, participants reported when they realized the first urge to move (W). All other instructions were kept the same.

In Libet's original experiment, the movement made by the participants was a flexion of their wrist or fingers (Libet, Gleason, et al., 1983, p. 625). We decided to use a mouse click, mainly because our technical setup did not allow us to control the computer using EMG

activation; hence it was impossible to stop the Libet's clock if no other input than EMG was provided. However, we assumed that this trade-off might bring some advantages. First, as a lot of recent Libet-style experiments employed a mouse click or a keypress as a substitute for the movement onset measured with EMG (e.g. Pockett & Miller, 2007; Soon et al., 2008; Verbaarschot et al., 2015; Caspar & Cleeremans, 2015; Dominik et al., 2017), we aimed to compare the time of a mouse click to an EMG onset. This was already done by Haggard and Eimer (1999) who found that the keypress typically occurs 30–50 ms after the EMG onset, and we aimed to verify their result. Second, the mouse click should represent a far more abrupt and better bounded movement than a flexion of a wrist or fingers. It should therefore be easier for the participants to determine the timing of their movement. Nevertheless, clicking a mouse button brings another problem that needs to be resolved. Banks and Isham (2009) showed that a delayed auditory feedback to the movement might systematically distort the W reports. Because the mouse click produces this auditory feedback (presumably delayed by 30–50 ms compared to the EMG onset, as Haggard & Eimer, 1999, suggest), it seems advisable to eliminate this factor. We managed that by putting soft earplugs into participants ears (after making sure that the participants do not have any question to be answered, of course). All participants reported that they were unable to hear the click and that they did not feel too uncomfortable while having the earplugs in their ears.

The **S series** also consisted of 40 or 41 trials for A and O modes of recall, respectively. Libet's reason to include this type of series was to have a "correction" for how accurate participants' time perception was (Libet, Gleason, et al., 1983, p. 627). In our case, participants were seated in the same position as in the M and W series, but as the event was a skin stimulus and not a mouse click, EMG was not recorded. Instead, a tactile stimulator was placed on the anterior side of participant's left wrist (see chapter 5.2.3), which was resting on the wooden board in front of the participant in supine position. The participant was instructed to sit calmly, fixate the mark in the centre of the clock and wait for the stimulus. The skin stimuli in the original experiment were delivered by an experimenter at random times unknown to the participant, but never during the first revolution of the clock (Libet et al., 1982, p. 325; Libet, Gleason, et al., 1983, p. 625). We used an algorithm running within our "Libet's clock", which sent an analog signal driving the stimulator at random times after the first revolution was completed. After the continuation interval following the stimulus, a response prompt was displayed asking about the time in which the participant

registered the stimulus (again, the prompt and the way of responding varied for the A and O modes of recall). As in the M and W series, participants were advised to relieve their eyes if needed, by looking at a wall before continuing the series.

The **P series** was similar to the M and W series in terms of experimental setting, but the task was different. Participants were sitting with their right hand on the mouse (EMG was recorded) and watched the centre of the clock. In this case, there was a fixed bright green dot on the clock's circumference in addition to the moving dot. This green dot—the target point—appeared on a random ("pre-set") position, which was different in each trial (as opposed to Libet's original design in which the target mark was placed on the same spot in each block of 10 successive trials). The participant's task was to click the mouse button when the moving dot reached the green mark's position, with maximum accuracy possible. After the click was made, no response prompt was displayed; instead, the target point disappeared, while the white dot continued moving for another few revolutions, after which a new target point was displayed at another location, so the task could be repeated. If a participant missed the target point, he or she was instructed to simply wait for the next revolution and then try again; thus, the clock did not force any pace on the participants. There were always 40 trials in the P series.

The **Pv series** was almost identical to the P series. The difference was that the participants were instructed to prepare to make the movement (click the mouse button) exactly at the time marked by the target point, but then stop ("veto") the movement just before it begun (for more details, see Appendix 5.8). If the participant believed that the task was done right, he or she clicked the mouse button at any time during the next revolution to give a signal that the veto was made. The target point then disappeared and reappeared in another location a few seconds later. If the participant was unsure that the veto was properly performed, he or she simply waited for another revolution without clicking and then tried again. As in the P series, there were always 40 trials in each Pv series.

The **Sp series** were not conducted in our replication because when we designed the study, we were unaware of the Sp series' existence since they were only mentioned in Libet, Wright, et al. (1983), which we missed in our research (see the introduction to chapter 2 for explanation of our mistake and its implications for our assertion about the concept of conscious veto). Even though I admit our error, I remain confident in our results. That is because omission of these series could not influence our replication data since Libet

conducted the trials with the Sp series within supplementary study realized after the main experiment was finished.

In our replication, the participants were initially trained in all types of conducted sessions. Besides the initial training sessions (see chapter 5.2.5), each series was preceded by 10 training trials consisting of the same task and employing the same mode of recall as the following 40 or 41 trials. This was done with respect to Libet's directions (Libet et al., 1982, p. 325) to re-familiarize the participants with the upcoming task. In Libet's original design, there was also an S series of 25 trials placed at the beginning of each regular session, which were also intended to train the participants—in this case to give accurate introspective reports. Libet did this by providing feedback on how close their responses were to the times of actual stimuli every 5 trials (Libet, Gleason, et al., 1983, p. 628). Additionally, Libet provided the feedback at the end of each regular S series (Libet, Gleason, et al., 1983, p. 627). At the beginning of chapter 3.3, I have discussed objections against this idea: (1) performing this feedback training over multiple sessions must inevitably lead to variable results because the participants would constantly improve in their accuracy in the S series (Gomes, 1998) and (2) the accuracy related to the S series cannot be considered a universal introspective bias, because it would change based on which modality is used for the stimulus (Danquah et al., 2008). With respect to these objections, we chose to withdraw from following Libet's directions in this case, deciding to refrain from providing any feedback in the S series, to remove the 25 initial S trials from all sessions in our design and not to correct the M and W reports based on the S reports. While being aware that this step might alter our results significantly compared to Libet's original outcomes, we believed that in this case it was more useful to conduct the experiment in a way, which was arguably more valid than replicatively accurate.

5.2.5 Progression of the experiment

As stated above, we conducted all the task variants, which were present in Libet's original design. We organized the series into 7 sessions. To emphasize the difference between the two distinctively different types of sessions, we introduced a new label *Supplementary session* for the sessions with the P and Pv series. In this chapter, I will introduce the content of each session and explain how the experimental conditions were rotated. See Table 4 for an overview of one possible experimental progress.

Session 1 *training*	Personal and medical history. EMG biofeedback. Spontaneous movement training. Explanation and training: M(A), W(O), S(A), S(O) series; 10 trials each.					
Session 2 *regular*	re-training W(A) (10 trials)	W(A) (40 trials)	re-training M(A) (10 trials)	M(A) (40 trials)	re-training S(A) (10 trials)	S(A) (40 trials)
Session 3 *regular*	re-training W(O) (10 trials)	W(O) (41 trials)	re-training M(O) (10 trials)	M(O) (41 trials)	re-training S(O) (10 trials)	S(O) (41 trials)
Session 4 *regular*	re-training M(A) (10 trials)	M(A) (40 trials)	re-training W(A) (10 trials)	W(A) (40 trials)	re-training S(A) (10 trials)	S(A) (40 trials)
Session 5 *regular*	re-training M(O) (10 trials)	M(O) (41 trials)	re-training W(O) (10 trials)	W(O) (41 trials)	re-training S(O) (10 trials)	S(O) (41 trials)
Session 6 *supplementary*	re-training P (10 trials)	P (40 trials)	re-training Pv (10 trials)	Pv (40 trials)	re-training W(A) (10 trials)	W(A) (40 trials)
Session 7 *supplementary*	re-training Pv (10 trials)	Pv (40 trials)	re-training P (10 trials)	P (40 trials)	re-training W(O) (10 trials)	W(O) (41 trials)

Table 4: An overview of a model experimental arrangement. The M, W, S, P and Pv indicate the series type, (A) and (O) indicate the absolute and order mode of recall, respectively. The grey fields mark the training session or the re-training 10-trials series.

The first session was a **training session**, which included taking personal and medical history, an EMG biofeedback, spontaneous movement training and familiarization with the M, W and S tasks including learning the difference between the A and O modes of recall. There was only one training session for each participant, as we found no reason to repeat the training for the second time.

The personal and medical history inquiry focused on potential contraindications of EEG and EMG recording or sources of EEG artefacts, such as epilepsy, psychopathology, head injuries or relevant medication. None of the participants reported any of the potential risks.

The purpose of the EMG biofeedback was to train the participants to click the mouse button abruptly enough so that the EMG activation reaches its maximum in the shortest time possible. This was done using the BIOPAC MP36 unit and the lesson template *L01 – Electromyography (EMG) I* in the BIOPAC STUDENTS LAB software, which contains two tasks suitable for our needs. The first of these tasks required recording and visualizing the EMG in a graph—this helped participants learn in real time what the recording looks like and how to click the mouse button so that the EMG activation is fast enough. The second task was to listen to headphones, in which the EMG signal was transformed into a form of a sound wave—this was intended to further help participants to make abrupt clicks, so that the rise in the sound volume was as short as possible. If the

experimenters present to the following sessions noticed that the EMG rise tended to be insufficiently steep, they notified the participant and asked him or her to make the click more abruptly.

After it was clear that the participant knew how to click "correctly" (which the participants typically achieved in 30–40 minutes), we presented Libet's clock to the participants. The experimenters explained its function and then presented four examples of tasks relevant to the following 4 regular sessions. The first presented task was M(A)—M series in absolute mode of recall—which we found the simplest to explain, followed by the W(O), S(A) and S(O) series. The participants were assured that the A and O modes of recall will not vary during a single session and that the changes in the present session are intended for instructional reasons only. During the M(A) and W(O) series, the participants were instructed to make the movement spontaneously, with no pre-planning, just as the urge appeared (this was in fact stated in all relevant textual instructions in the following sessions, see the Appendix 5).

Also, the technical equipment was shown, described and demonstratively used during the training session, so that all participants were familiar with as many aspects of the experiment as possible. Special care was given to whether the participants feel comfortable with the equipment (i.e. the EEG cap, the EMG electrodes, the skin stimulator or the earplugs). However, no explicit statement about Libet's results or conclusions was made until the end of the last session.

The **regular sessions** consisted of three 10 trials re-training series and three regular series in the M, W and S conditions. Before each of the series, a separate textual instruction was displayed, and the participants were encouraged to read it every time to remind them of the current task and of some principles, which were constant during the experiment (such as the blinking rules, fixation mark, spontaneity of the movement etc.). In the first several sessions, the participants were required to repeat the instruction in their own words, so that it was clear that they understand.[5] After verifying that the participant comprehends the task, the first series (either M or W) begun. After the first 100 or 102 trials (10 re-training W, 40 or 41 regular W, 10 re-training M, 40 or 41 regular M) were completed, two questions were asked:

---

[5] Once the experimenters were sure that a participant has a grasp of the various variants of the procedure, the repetition of the instructions was no longer required.

1. "Did you notice during the experiment that you felt that the movement was not spontaneous, that you pre-planned it?"
2. "Did it occur to you during the experiment that you were surprised by the movement? That it came on its own, without you knowing?"

If the participant answered any of these questions positively, he or she was asked in which series it was, and the answer was recorded in the protocol (see below).

Once the M and W series were finished, a change in the technical setting was made to prepare for the S series (see chapter 5.2.3). Then, another 50 or 51 trials with skin stimulation were conducted (10 re-training S, 40 or 41 regular S).

The sixth and seventh sessions were the **supplementary sessions** and differed from the regular ones by introducing the P and Pv series. At the beginning of the sixth session, the participants were given training in the P and Pv tasks, which was similar to the training of the M, W and S series in the initial training session. Once the participants were comfortable with the new task, they were presented with three series: P, Pv and a supplementary W series. The W series in the supplementary sessions did not differ from the corresponding series in the previous sessions—its inclusion was intended to increase the number of the W trials, as these seem to be the most crucial aspect of Libet's experiment. This scenario was repeated in the seventh session, but with reversed order of the P and Pv series and opposite mode of recall for the W series.

The experimental conditions in the regular and supplementary sessions were rotated based on several principles:

1. The mode of recall stayed constant during each session.
2. The mode of recall alternated in consecutive sessions (both regular and supplementary).
3. The order of the M and W series was reversed between the 3rd and the 4th session.[6]
4. The order of the P and Pv series was reversed between the 6th and the 7th session.
5. The order of the P and Pv series was independent of the mode of recall in the following W series.

---

[6] This was not fully achieved in one case due to experimenters' error. See the asterisk in Table 5.

6. Every possible combination was assigned to two participants (one in the first group and one in the second group).

These principles resulted in a schedule depicted in Table 5.

	Participant 1 and 5	Participant 2 and 6	Participant 3 and 7*	Participant 4 and 8
Session 1 *training*	EMG biof. – M(O) – W(A) – S(A) – S(O)	EMG biof. – M(O) – W(A) – S(A) – S(O)	EMG biof. – M(O) – W(A) – S(A) – S(O)	EMG biof. – M(O) – W(A) – S(A) – S(O)
Session 2 *regular*	W(A) – M(A) – S(A)	W(O) – M(O) – S(O)	M(A) – W(A) – S(A)	M(O) – W(O) – S(O)
Session 3 *regular*	W(O) – M(O) – S(O)	W(A) – M(A) – S(A)	M(O) – W(O) – S(O)	M(A) – W(A) – S(A)
Session 4 *regular*	M(A) – W(A) – S(A)	M(O) – W(O) – S(O)	W(A) – M(A) – S(A)	W(O) – M(O) – S(O)
Session 5 *regular*	M(O) – W(O) – S(O)	M(A) – W(A) – S(A)	W(O) – M(O) – S(O)	W(A) – M(A) – S(A)
Session 6 *supplementary*	Pv – P – W(A)	Pv – P – W(O)	P – Pv – W(A)	P – Pv – W(O)
Session 7 *supplementary*	P – Pv – W(O)	P – Pv – W(A)	Pv – P – W(O)	Pv – P – W(A)

Table 5: An overview of the rotation of experimental conditions. The participants 1, 2, 3 and 4 were members of the first group studied in the first two months; the participants 5, 6, 7 and 8 were studied in the following two months. The M, W, S, P and Pv indicate the series type, (A) and (O) indicate the absolute and order mode of recall, respectively. The asterisk indicates that in the case of the 7th participant an error occurred resulting in the order of the sessions being altered (namely, the 3rd and the 5th sessions were swapped).

Since two experimenters out of five attended every session, a standardization of the procedure was needed. Therefore, we employed textual **protocols**—documents containing step-by-step directions on how to conduct each session, describing the most specific details, such as which connectors should be placed into which socket or which information should be entered in Libet's clock settings. The protocols were participant- and session-specific, which means that no pair of the 56 protocols was identical. Each protocol also contained several blank fields for procedural and technical notes and participants' answers to the standardized questions (see above).

5.2.6 Data analyses

The data were acquired from multiple sources in different forms and needed to be unified in a single framework. We aimed to use the R software for the analyses and therefore needed all the data converted into compatible formats.

The introspective and technical data from Libet's clock (i.e. the subjective reports of the M, W and S times, the timing of the skin stimuli in the S series and the data on when a click occurred in the M, W, P and Pv series) were exported as CSV files from the database, to which the Libet's clock software had sent all the data. The physiological recordings (one EMG and six EEG channels) and other data obtained using the MP150 unit (the keypress channel and the stimuli channel) were exported as CSV files from the AcqKnowledge software. These two types of CSV files were merged and temporally aligned based on the mouse clicks or the skin stimuli, which were recorded by both Libet's clock and the MP150 unit (see chapter 5.2.3). There were two merged files for each regular session (one for the M and W series, one for the S series) and one merged file for each supplementary session (including all the P, Pv and W series). The merged files were then segmented into epochs spreading from 2500 ms before the event (mouse click or skin stimulus) to 800 ms after the event.

The EMG data were analysed to find the onset of the EMG activation (which I will call the **EMG$_0$**), which serves as a reference point for temporal analyses of the M, W and P series. The EMG data was processed in three steps. In the first step, we calculated the absolute value of each EMG waveform. In the second step, the Butterworth procedure was applied to rectify the data. The third step was to find the EMG onset, which we did by calculating a threshold value equal to five times the IQR (interquartile range) of a segment from -1500 to -300 ms added to the respective Q3 (upper quartile); this threshold was subsequently applied to the segment of -300 to +200 ms (we extended this range into the positive values, because in some cases in the P series the participants were expected to exceed the zero time, which in this case was marked by the position of the target point).

The EEG data were treated separately for each channel. Before any further analyses, we rejected those EEG recordings, which contained excessive noise (284 recordings in 68 series across all channels, sessions and participants, out of 858 recordings in total). The remaining recordings were additionally filtered (using 0.5 Hz high-pass and 35 Hz low-pass filter) to reduce or eliminate long polarizing shifts of the EEG baseline and the residues of the 50 Hz

noise. These filtered recordings were segmented into 34656 epochs where every epoch represented one trial and one electrode. Epochs from the M, W and P series containing no EMG onset or an invalid EMG onset and epochs from Pv series containing an EMG onset were rejected (5640). Epochs, which contained blinking artefacts and artefacts caused by eye movement were further discarded if the voltage P-P (peak-to-peak) range exceeded a rejection value of 0.15 mV (11347).

Every remaining epoch (17669) was then temporally locked to the respective **reference point**: the EMG onset in the M and W trials, the skin stimulus in the S trials and the target point position in the P and Pv series.[7] Each channel in each series was then averaged. Averages calculated from less than 10 trials (22) were further rejected. The averaged data line plots were used to represent the mean course of activation on a given electrode before and after the reference point. Before further analyses, grand averages were calculated (i.e. we plotted the mean courses of activation based on all 17669 valid trials in respective series regardless of the session, but with respect to the electrode). To assess the RP onsets in individual series, we used both methods suggested by Libet, Gleason, et al. (1983, p. 632): (1) the MN (main negative) method and (2) the RP$_{90\%}$ method.

The **MN method** consists of an eye-ball inspection of each graph by multiple independent investigators. For our analyses, we employed five members of our team, who were all aware, what the onset should look like (we designed a custom "Guide to the RP identification", see the Appendix 6), but who were not aware of the type of series a respective graph was obtained in. The only exception was myself, because my task was also to merge the estimates into one file and, for that reason, I was aware of the respective series types at the time. The conclusions were then confirmed by one independent examiner who was not a member of the authors' team. While estimating whether an RP is present, we proceeded rather conservatively (i.e. we did not identify a negativity as an RP if the waveform lacked some important characteristics, such as adequate artefact-free straight baseline).

The **RP$_{90\%}$ method** is a more objective approach to assessing the RP onset. It uses a calculation of the area under the RP curve preceding the reference point (see Libet, Gleason,

---

[7] This is because in the Pv series, an EMG onset is not supposed to be found (as the participant is told to veto the movement). Therefore, we analyse the EEG and calculate the RP onsets in the Pv series relative to the target point. We performed the same procedure with the P series to make the RP onset times in the P series comparable to those in the Pv series.

et al., 1983, pp. 632–633). Unlike Libet, we did not have to take geometric measurements from a paper tape, as we used a computer algorithm based on Libet's calculations instead. The algorithm was applied to the averaged EEG plots. A baseline was calculated for each plot as an average of the EEG waveform in the interval -1500 to -1000 ms in the M, W and S series and -2500 to -1500 ms in the P and Pv series.[8] A window 50 ms wide started with its right edge aligned with the reference point and moved to the left (i.e. to the negative values) by steps of 1 ms. In each step, the area under the EEG waveform (or rather above it, as we plotted negative potentials below the baseline) was calculated. If the window entailed the curve both under and above baseline, the upper portion of the area was subtracted from its lower portion. Once the area under the curve was smaller than 12.5 µV/50 ms, all remaining data points preceding the current window position including the segment currently covered by the window were considered 0. The RP onset was placed to the lower limit of 90% portion of the area (calculated from positive to negative, starting from the reference point). Our algorithm was in fact more precise than Libet's original calculation, as it used a window of constant size.

Three researchers also assessed the RP types (I, II or III). Libet et al. (1982, p. 326) describe **type I RP** as follows: *"In type I RP a gradually or steadily rising, ramp-like form begins distinctly prior to -700 msec"*; the authors also add that more extreme examples of type I RP occurred in early sessions before the spontaneity requirement was added into the instruction. Because our instruction contained a mention of spontaneity in all sessions, we expect type I RPs to be rather rare and if it occurs, we expect it to begin later. The ideal depiction of type I RP is shown in the Figure 3a in chapter 2.4.1.

Type II RP is described as follows: *"In type II RPs, the main rise of negativity starts in the range of about -400 to -700 msec, (...) The main portion of this RP is often somewhat dome-shaped rather than ramp-like in form"* (Libet et al., 1982, p. 326). The authors admit that the RP II may be preceded by some irregular negativity but suggest that this negativity does not have the characteristics of the early rise of type I RP. The ideal type II RP is depicted in Figure 3b in chapter 2.4.1.

---

[8] Later baseline (-1500 to -1000) was preferred, because gradual increases and decreases in voltage early in the baseline might shift the RP onsets; however, in the P and Pv series, the RPs are expected to have earlier onsets than those in the M and W series; therefore, the baseline was moved to -2500 to -1500 in these series.

Type III RP seems to be less frequent potential characterized as follows: "*In type III RPs, the main rise of negativity does not appear until about -250 to -200 msec (...). Total durations of any detectable negativity and especially total areas of RP are also low*" (Libet et al., 1982, p. 326). For an image of an ideal type III RP, see Figure 3c in chapter 2.4.1. In our analyses we extended the type III RP definition to any late-onset negativities that cannot be regarded type II RPs (some of our type III RPs' onsets suggested type II RP, but their shape was rather sharp, not dome-shaped—these were thus identified as type III RP).

Libet also suggested that the pre-set series elicit a special ramp-like RP with an onset earlier than -1400 ms (which was the edge of Libet's epoch range; see Libet et al., 1982, p. 330). We adopted this view, but had to extend the pre-set RP definition, because RPs in our pre-set series exhibited rather dome-shaped form (see chapter 5.3.3).

Additionally, four of the examiners also assessed whether a P300 waveform is present, as it is expected to follow the stimulus delivery in the S series (see Libet et al., 1982, pp. 330–331; for the P300 characteristics, see e.g. Picton, 1992). The actual occurrence was then finally judged by me within the judgment-merging procedure (see above in this chapter).

The M and W data obtained using Libet's clock in the A mode of recall were analysed for each trial by subtracting the EMG_0 timestamp from the timestamp of participant's response. For example, if a participant made a W report at a location which corresponded to 204 ms before the EMG_0, that single W time would be -204 ms. The S reports made in the A mode of recall were analysed similarly; the timestamp of the skin stimulus was subtracted from the timestamp of the participant's response.

The responses acquired in the O mode of recall could not be analysed for each trial and were calculated for the whole series using the formula presented at the end of chapter 5.2.2. The M and W reports were then corrected for the mean EMG_0 calculated from the whole respective series.

In both cases, the W reports could also be corrected by the mean S report, but as explained in chapter 5.2.4, we decided to refrain from such procedure. However, as we published the introspective reports data (see Dominik et al., 2018b), I invite interested researchers to explore this idea on our dataset.

Other analyses presented in the chapter 5.3 either use standard statistical procedures briefly described along the relevant results.

5.3 Results

5.3.1 EMG onset timing

The EMG onset is relevant for the M, W and P series. In the M and W series, it constitutes the reference point for the EEG onset and the introspective reports. In the P series, we are interested in the comparison of the target point, the mouse click and the EMG onset.

EMG onsets in M and W trials

The M and W series serve specific interpretational purpose, and thus we treated the EMG onset in these series separately from the P series. There were 3239 individual trials in 80 M or W series regardless of the mode of recall. In 271 trials, no EMG onset was found. Another 98 trials contained EMG onsets preceding the mouse click by more than 150 ms. Additionally, 169 EMG onsets occurred after the mouse click by 1 to 106 ms. We discarded all 271 trials with no EMG onset found, 98 trials with EMG earlier than 150 ms before the click and 169 trials, in which the EMG onset followed the click. We were thus left with 2701 valid EMG onsets in the range from -150 ms to 0 ms (compared to the mouse click), which were included in the RP analyses and the analyses of introspective reports.

The mean difference between the EMG onset and the mouse click in the M and W series with no trials discarded is -62.5 ms (SD = 42.8). After discarding the EMG onsets earlier than -150 ms and later than 0 ms, the mean value practically did not change (M = -62.4 ms; SD = 19.2). Haggard and Eimer (1999) assert that the typical EMG onset occurs 30–50 ms before the mouse click. In our case, only 567 of the 2701 valid results were within this range. Instead, the typical majority (2441 trials, 90 % of the valid trials in range from -150 to 0 ms, 83 % of all trials) occurred between -90 (excluding) and -30 ms (including). For more synoptic overview, see Figure 8. Using the Wilcoxon signed-rank test, we found that in the case of the M and W trials the EMG onsets differ significantly from the mouse click, $Z = 45.459$, $p < 0.001$, but we also claim that the EMG onset may commonly precede the press of a button by more than Haggard and Eimer (1999) suggested.
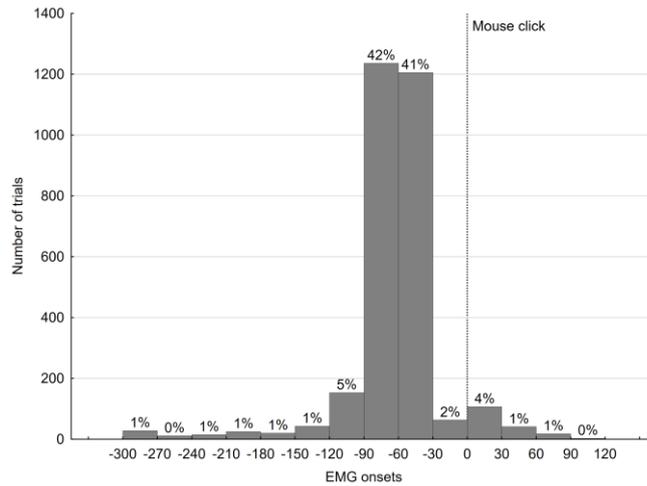
Figure 8: The histogram of EMG onsets in the M and W series compared to the mouse click (the x-axis values signify the intervals' upper boundaries; trials with no EMG onset found discarded; n = 2968).

EMG onsets in P trials

The EMG onsets in the P series (together with the mouse click times) show how accurate the participants were while attempting to click at the time marked by the static target point on Libet's clock. We are interested in three kinds of time differences: (1) between the EMG onset and the mouse click (as in the case of M and W series), (2) between the EMG onset and the target point and (3) between the mouse click and the target point. There were 640 trials in the P series (we do not include the Pv series). In 15 trials, no EMG onset was found. Sixteen EMG onsets occurred more than 150 ms before the mouse click, 35 EMG onsets occurred after the click. We did not discard any trials from the RP analyses—except for those with no EMG onset found—as we relate the RP onsets to the target points and not the movement onset. The mean of the EMG onsets relative to the mouse clicks including all trials was -64.6 ms (SD = 41.9); if we include only the EMG onsets later than -150 ms, but earlier than 0 ms, the mean value is -66.6 ms (SD = 21.1). For an overview of the difference between the EMG onset and the mouse click in the P series, see Figure 9a.

As for the participants' accuracy in the P series measured by the EMG, the mean EMG onset occurred 35.4 ms before the target point (SD = 64.4, see Figure 9b). If we measure the accuracy by the mouse clicks, the mean mouse click occurred 28.4 ms after the target point (SD = 67.4, see Figure 9c). Overall, we can state that the participants are generally accurate in clicking at the pre-set target time, with the EMG onset preceding the target time and the actual mouse click being slightly delayed.
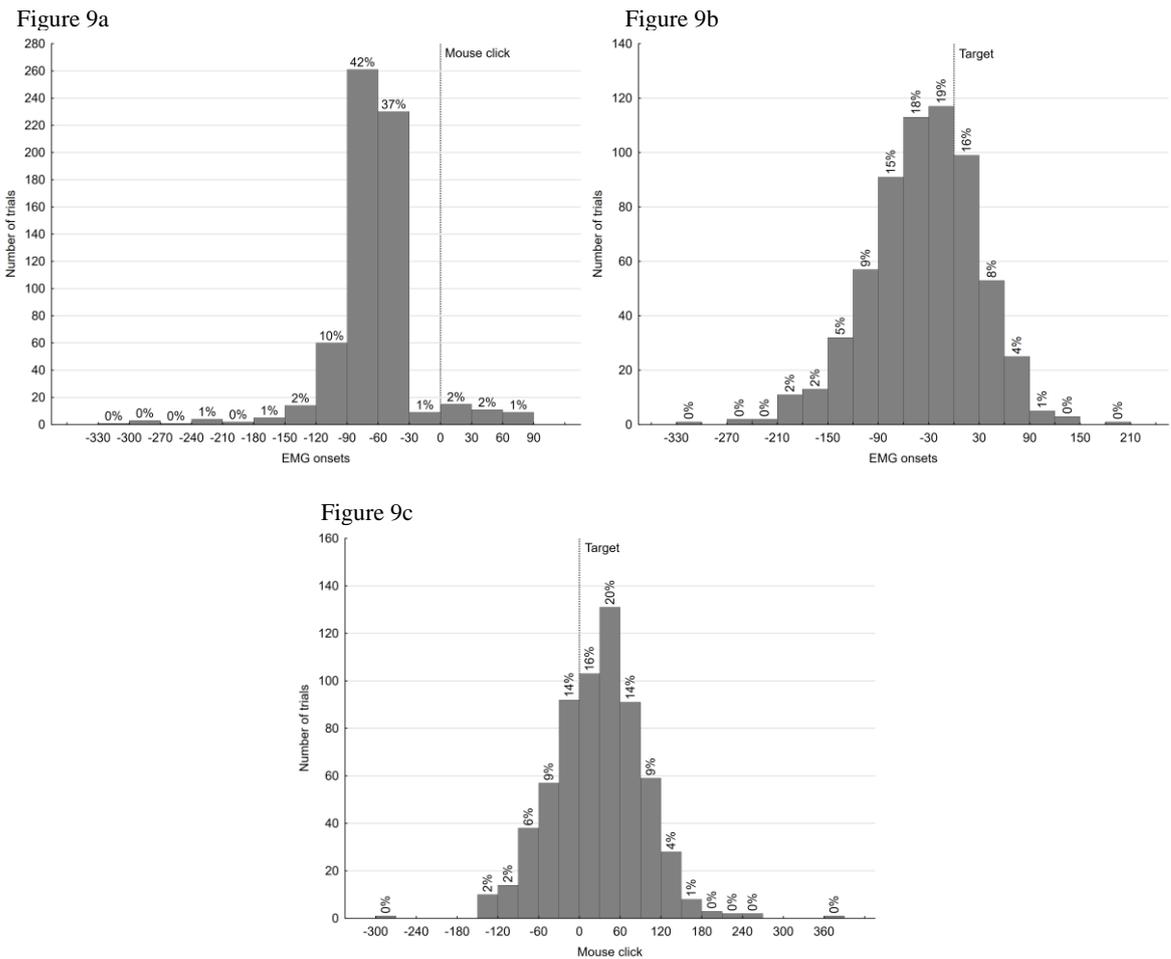
Figure 9: The histograms of EMG onsets in the P series compared to the mouse click (n = 625, trials with no EMG onset found discarded, Figure 9a), EMG onsets in the P series compared to the target point (n = 625, trials with no EMG onset found discarded, Figure 9b) and the mouse clicks in the P series compared to the target point (n = 640, no trials discarded, Figure 9c). The x-axis values signify the intervals' upper boundaries.

5.3.2 Introspective impressions timing

All introspective reports (M, W and S) were treated separately for the absolute (A) and order (O) mode of recall (see chapter 5.2.2). All temporal data in this section are related to their respective EMG onset or stimulus delivery (in the O mode of recall, the mean value is related to the mean EMG onset or the stimulus delivery calculated for the whole series).

Absolute introspective reports

The absolute reports of the spontaneous movement initiation, i.e. the M(A) reports, were collected in fifteen 40-trials series (originally, 16 series were carried out, but one M(A) series with participant 4 was corrupted due to a technical error). Only trials with a valid EMG onset in the range from -150 to 0 ms (531 out of the 600 trials) were included in the analyses. The

grand averaged M time was reported slightly after the EMG onset with substantial amount of variability (M = 26.7 ms, SD = 134.7, see Figure 10a). The M report mean estimate based on the whole series (i.e. average of the mean reports) is M = 30.9 ms (SEM = 79.8).

The absolute reports of the first urge or wanting to move, i.e. the W(A) reports, were collected in twenty-four series with 808 out of the 960 trials containing valid EMG onset. The overall mean W time was reported about 100 ms before the EMG onset with even larger variability than in the case of the M reports (M = -98.5 ms, SD = 197.6, see Figure 10b). Calculating the mean M report using the series means: M = 101.2 ms (SEM = 151.1).

Figure 10a

Figure 10b
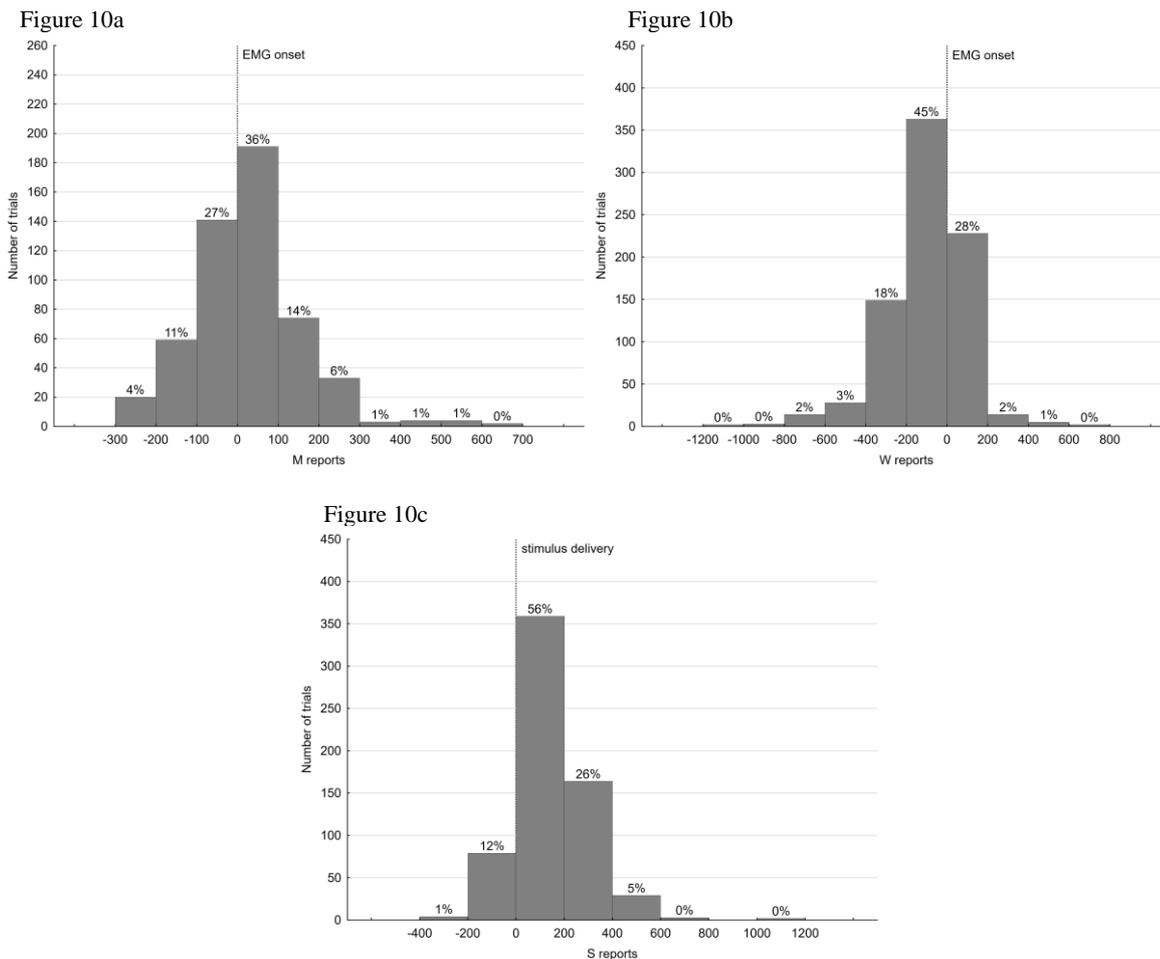
Figure 10c

Figure 10: The histograms of the M(A) reports (n = 531, Figure 10a) and W(A) reports (n = 808, Figure 10b) compared to the EMG onset and S(A) reports (n = 640, Figure 10c) compared to the skin stimulus delivery.

The absolute reports of the skin stimulus registration, i.e. the S(A) reports, were collected in sixteen 40-trials series with no trials discarded (the EMG recordings would be irrelevant in this case and the skin stimulator did not exhibit any significant unsystematic error, therefore no trials had to be rejected). The analysis of the 640 S(A) trials showed that the skin stimulus was reported to be registered on average about 150 ms after the stimulus

(M = 146.0 ms, SD = 150.2, see Figure 10c); calculating using the series means: M = 146.0 ms (SEM = 77.3).

Figure 11 contains an overview of the means and confidence intervals of the M(A), W(A) and S(A) reports for individual participants. The graph suggests two intriguing findings regarding the W reports: first, that three participants reported the W values on average after the EMG onset, and second, that participant 3 tended to state similar reports regardless of the M and W instructions. These findings are further discussed in chapter 5.4.4.
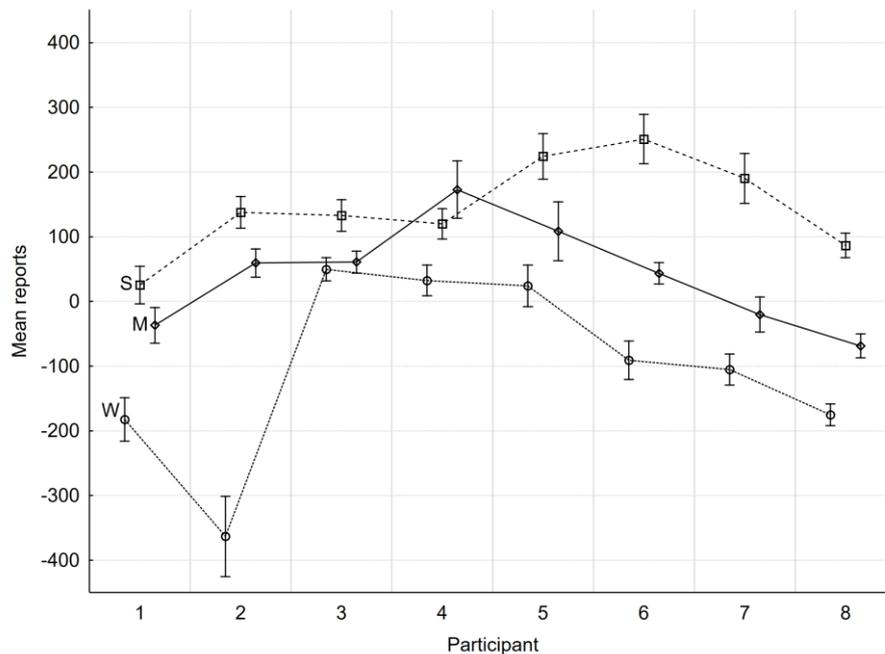
Figure 11: A graphical representation of the reports' means based on individual participants (x-axis) and the series type (separate lines). Note that the participants' order is arbitrary and that the slope of the connecting lines does not provide any information. The vertical bars denote the 95% confidence intervals. Value 0 on the y-axis denotes the EMG onset or the skin stimulus delivery.

It is reasonable to ask whether the absolute reports differ based on the series type (M, W, S). To answer the question, we employed a mixed-effect linear model testing the effect of the series types (**Series**) on the reported value, with following covariates: **Subject** (a random categorical factor describing the differences between the individual participants), **Trial** (a fixed continuous factor describing the succession of individual trials in the series) and **Session** (a fixed continuous factor describing the differences between individual sessions). The Trial and Session factors were centred by subtracting the mean. The covariates are all included mostly to consider the fact that the individual trials are not independent (every 40 trials in a single series came from the same participant, series of the same type were always

parts of different sessions and the trials might have brought different reports based on their succession in the respective series). We also included the participant-series interaction term (**Subject*Series**; random factor) as a covariate to take into account the possibility that the participants may provide different reports in specific series. We excluded 15 trials as residual outliers. The model exhibits a satisfying fit and explains 54.8% of the reports' variability (30.8% if the random factors Participant and Participant*Series are excluded). For an overview of the effects, see Table 6.

Factor	Estimate	Std. error	t value (df)	p value
Intercept	41.13	25.20	1.632 (7.90)	0.142
Series (S)	106.09	26.90	3.944 (7.80)	0.004**
Series (W)	-146.49	37.23	-3.934 (8.00)	0.004**
Session	9.26	2.21	4.194 (1950.80)	< 0.001***
Trial	0.81	0.24	3.316 (1940.90)	< 0.001***

Table 6: An overview of individual effects in the mixed-effect model analysing the M, W and S reports collected within the absolute mode of recall. The random factors are included, but not displayed. The factors Series (S) and Series (W) are referenced to mean M reports.

The main effect of the factor Series is highly significant, $\chi^2(2) = 13.804$, p = 0.001. These results suggest that there is an overall difference between the M, W and S reports. Specifically, the W reports are significantly earlier than the M reports, t(8.00) = 3.934, p = 0.004, the M reports are significantly earlier than the S reports, t(7.80) = 3.944, p = 0.004, and the W reports are significantly earlier than the S reports, t(8.00) = 5.968, p < 0.001. For the estimates comparison of the M(A), W(A) and S(A) reports, see Figure 12.
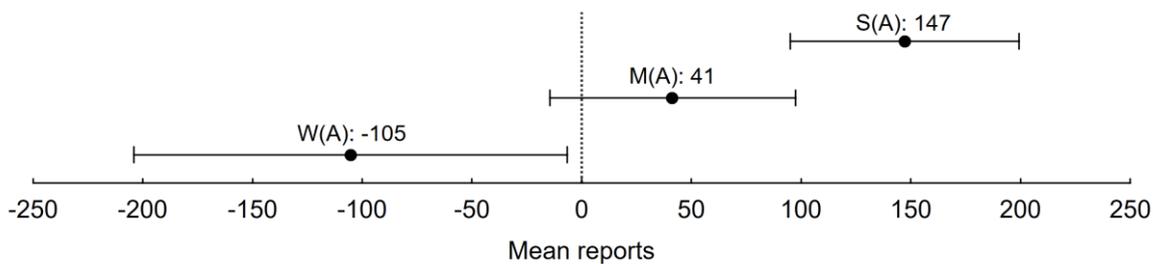
Figure 12: A comparison of the estimates of the W(A), M(A) and S(A) reports. The horizontal bars denote the 95% confidence intervals. Value 0 on the x-axis denotes the EMG onset for the M and W reports and skin stimulus delivery for the S reports. Note that these results do not correspond precisely to the means for the absolute reports presented throughout chapter 5.3.2 due to the inclusion of covariates' effects in the present estimation.

Additionally, we tested whether the W reports change when the W task is carried out before or after the M task (see Study 1 of this thesis). We analysed the **Order** effect (W-M or M-W) on the W reports (**Trial** included as a fixed covariate, combination of **Subject** and **Session** as a random covariate). We found that the Order has no significant effect on the W reports, $t(16) = 0.162$, $p = 0.873$, although the W reports acquired after the M task were indeed earlier (M = -115.9, 95% CI [-224.4, -7.4]) than the W reports acquired before the M task (M = -103.9, 95% CI [-212.3, 4.4]).

Order introspective reports

The reports made in the trials with the order (O) mode of recall can be calculated as means for the whole series but not the individual trials. Therefore, we cannot calculate the SD for the series, but we can calculate the SEM (standard error of the mean) as the standard deviation of the mean estimates. Based on that calculation, we can also assess the 95% confidence intervals for the mean based on the SEM and the number of series (not the number of trials as in the case of the A reports).

The mean order report of the movement initiation, i.e. the M(O) report, was remarkably close to the EMG onset (M = 0.0, SEM = 83.5). The mean W(O) report preceded the EMG onset by about 70 ms (M = -70.3, SEM = 104.7). The S(O) reports followed the skin stimulus slightly (M = 34.1, SEM = 69.4). Because the observed values are means and not individual reports, we cannot conduct the same regression procedure as in the case of the absolute reports and make the estimates of the M(O), W(O) and S(O) means after taking the participants, session and trial into account. For an overview of the actual means and the confidence intervals, see Figure 13.
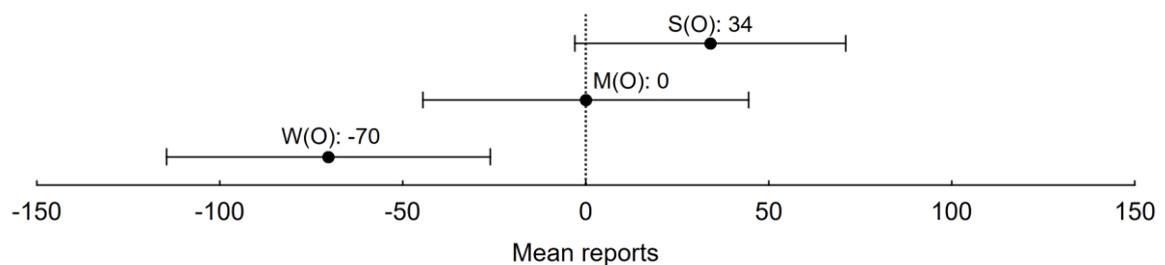

Figure 13: A comparison of the mean W(O), M(O) and S(O) reports. The horizontal bars denote the 95% confidence intervals. Value 0 on the x-axis denotes the EMG onset for the M and W reports and skin stimulus delivery for the S reports.

Finally, we also present the comparison of the A and O modes of recall. Because the Figures 12 and 13 present different approaches to the mean estimation, we decided to calculate the means and the confidence intervals for the A mode of recall the same way as

we did in the case of the O reports (i.e. for the series means, not the individual trials). The resulting estimates are presented in Figure 14. The figure shows that the order reports tend to be less extreme (i.e. closer to the EMG onset or the stimulus delivery). Inspection of the 95% confidence intervals suggests that there is a significant difference between the S reports made in the opposite modes of recall; other types of reports do not differ significantly based on the mode of recall.
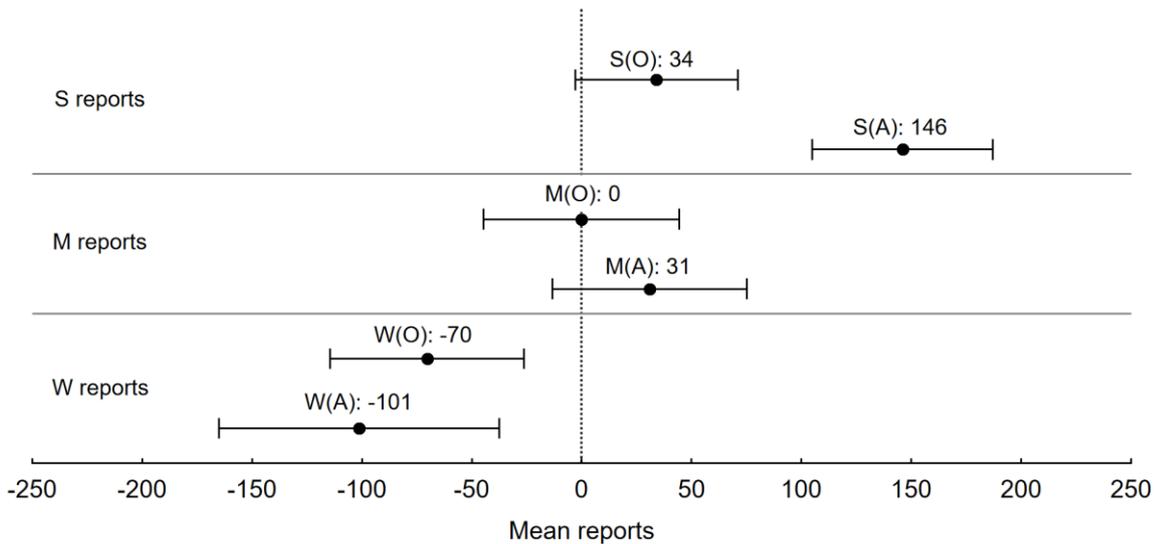
Figure 14: A comparison of the mean W, M and S reports made in the absolute (A) and order (O) modes of recall. The horizontal bars denote the 95% confidence intervals. Value 0 on the y-axis denotes the EMG onset for the M and W reports and skin stimulus delivery for the S reports.

5.3.3 Event-related potentials occurrence and timing

We examined two types of event-related potentials (ERPs for short): the readiness potential (RP) and the P300 wave (see chapter 5.2.6). The RP is expected to precede the movement onset in the M, W, P and Pv series and their presence and onsets were assessed using two methods (the eye-ball MN method and the computational $RP_{90\%}$ method). The P300 is a cognitive ERP expected to follow the stimulus presentation in the S series.

Contrary to our original expectations, the EEG recordings contained a large amount of noise, which was true even after averaging the signal. Even though this may limit the validity of our findings, our data still demonstrate many interesting points.

Most importantly, the grand averages in the corresponding series types (i.e. the M/W, P, Pv and S) contain the expected waveforms distinctly identifiable on the C_z, C_3, C_4 and P_3

electrodes (see Figure 15).[9] The Fp$_1$ and Fp$_2$ electrodes recordings were often distorted and did not exhibit the expected potentials; these were thus excluded from the grand average plots. Table 7 shows how many trials were included into each grand average calculation.

	C$_z$	P$_3$	C$_3$	C$_4$
M/W	1759	1818	1672	1747
S	464	835	776	792
P	197	197	194	198
Pv	498	497	476	469

Table 7: Number of trials included into each grand-averaged ERP plot. The M and W trials are in this case considered to be equivalent.

Figure 15a shows the general readiness potential negativity with the beginning preceding the EMG onset substantially. The grand-averaged RP has the largest amplitude on the C$_z$, followed by the C$_3$ (contralateral to the moving hand) and C$_4$ (ipsilateral to the moving hand). The grand-averaged RP on the P$_3$ electrode is irregular and exhibits a positive spike near the negativity peak. The grand-averaged RP on the C$_z$ resembles a ramp-like shape characteristic for the type I RP, but it begins later than the type I RP is expected to. On the C$_3$ and C$_4$, the grand-averaged RP resembles a dome-shaped type II RP.

Figure 15b demonstrates clearly two important points. First, no negativity with the RP characteristics precedes the stimulus onset. Second, the stimulus delivery is followed by a wave complex ending in a large positive spike, which we identified as the P300 positivity.

Figure 15c shows that when a subject is instructed to make the movement at the target time, the movement is preceded by large negativity preceding the movement by about 1500 ms. This negative potential does not exhibit typical characteristics for any of the RP types—it is dome-shaped, but begins more than 700 ms before the movement, so it should not be regarded type II RP; it also does not fit the pre-set RP description by Libet et al. (1982, p. 330), who suggested that the pre-set RP should exhibit a ramp-like form. Therefore, as stated in chapter 5.2.6, we decided to extend the pre-set RP definition to any early onset negativity preceding the movement during the P and Pv series.

---

[9] All the ERP plots presented here depict the positive charge above the baseline. I am aware that this disagrees with some older conventions, but I find it easier to communicate the results this way when reporting both negative and positive potentials.

Figure 15d shows a negativity more closely fitting Libet's definition of the pre-set RP. The potential precedes the target time (and the moment of conscious veto, respectively) by about 1500 ms and has a ramp-like growth.

Figure 15a
M and W series (grand-averaged as one series type)

Figure 15b
S series (grand-averaged)

Figure 15c
P series (grand-averaged)

Figure 15d
Pv series (grand-averaged)
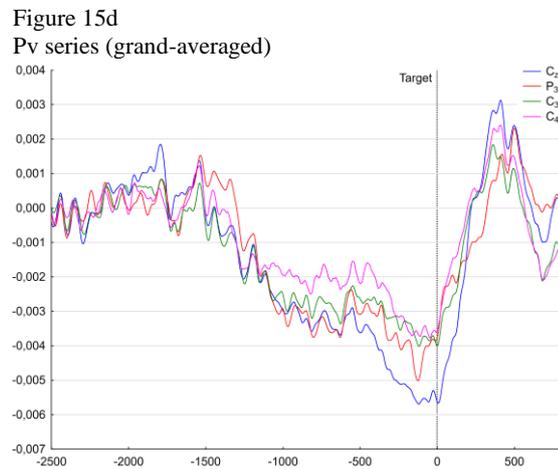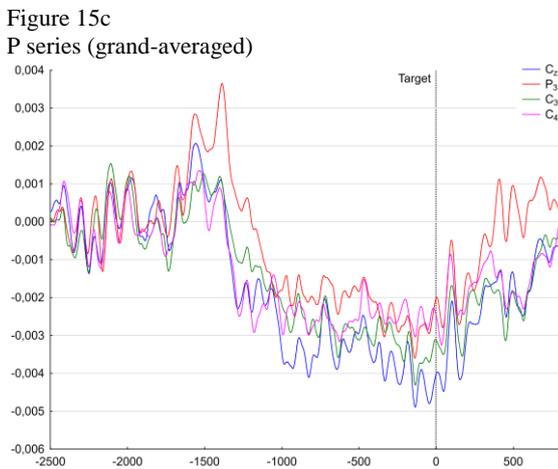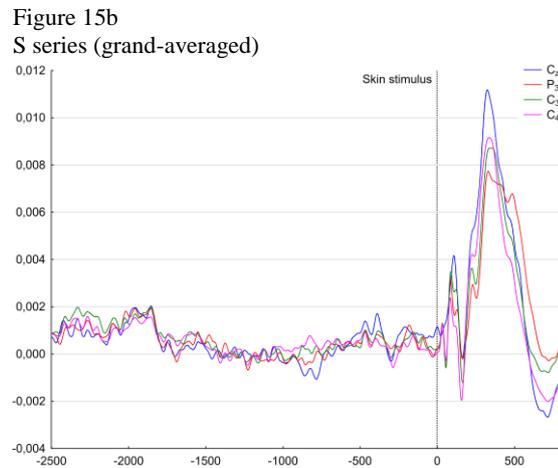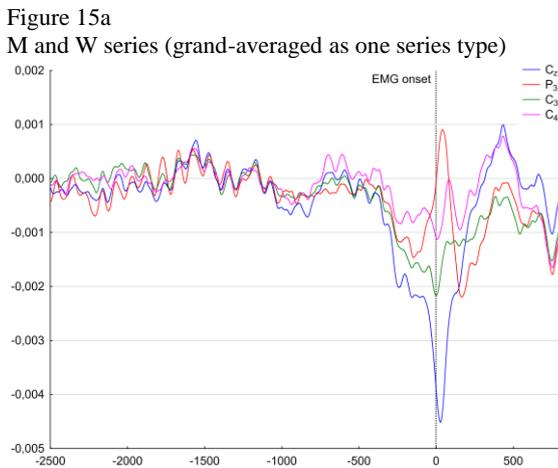
Figure 15: The grand-averaged ERPs in the four series types. Figure 15a shows distinct RP negativity beginning prior to the EMG onset. Figure 15b shows a P300 positivity following the stimulus. Figure 15c shows irregular dome-shaped negativity beginning long before the pre-set target point on Libet's clock was reached in the P series. Figure 15d demonstrates ramp-like negativity beginning long before the target point was reached in the Pv series, which in this case should be the time when the vetoed movement was supposed to be realized. The baseline in Figures 15a and 15b was estimated as the mean voltage in the time interval <-1500;-1000>; the baseline in Figure 15c and 15d is estimated as the mean voltage in the time interval <-2500;-2000> (because the observed potentials in the P and Pv series are longer).

These results show that although the recordings were noisy, the outcomes seem to be valid. I will now investigate the individual series types and the respective ERPs found in them. Table 8 lists which RP types were identified on C_z in which series (I consider the C_z electrode to be the most informative). I only list the series in which a readiness potential on the C_z electrode was found using the MN method. The RP$_{90\%}$ method seemed to be more

sensitive to RPs obscured by the noise hindering the MN detection; however, as we aimed to classify all RPs as type I, II, III or pre-set, we used the $RP_{90\%}$ as a supplementary method, because if the RP was not identified using the MN method, the RP type could not have been assessed. Because of the low frequency of identified RPs, we cannot perform any inferential statistical procedure and will thus report descriptive statistics only.

	RP I	RP II	RP III	pre-set RP	no RP found	P300	valid series
M	**3 (15 %)**	**4 (20 %)**	**2 (10 %)**	0 (0 %)	11 (55 %)	0	20
W	**6 (19 %)**	**6 (19 %)**	**4 (13 %)**	0 (0 %)	16 (50 %)	4	32
P	1 (8 %)	2 (15 %)	0 (0 %)	**7 (54 %)**	3 (23 %)	2	13
Pv	0 (0 %)	0 (0 %)	0 (0 %)	**6 (46 %)**	7 (54%)	0	13
S	0 (0 %)	0 (0 %)	0 (0 %)	1 (8 %)	**11 (92 %)**	8	12

Table 8: Number of ERPs found on the C_z electrode in individual series. The bold frequencies mark which ERPs are expected to occur in the respective series. The percentages show the relative frequency of the respective RP type compared to the number of valid series of the given kind. The percentages are not stated in the P300 column, because the P300 did not in our procedure rule out the RP occurrence (the percentage would therefore be confusing as their row sum would exceed 100 %).

ERPs in the M and W series

Altogether, we analysed 303 plots with valid EEG acquired in the M or W series, including all 6 electrodes. We found **21 ramp-like type I RPs** in 14 series (9 on C_z, 2 on C_3, 5 on C_4, 3 on P_3, 1 on Fp_1 and 1 on Fp_2) across 11 sessions. The average type I RPs' onset was estimated to be -667.4 ms (SD = 168.7) using the MN method and -580.3 ms (SD = 185.3) using the $RP_{90\%}$ method. Figure 16a demonstrates typical type I RP.

Further, we found **37 dome-shaped type II RPs** in 18 series (10 on C_z, 9 on C_3, 5 on C_4, 9 on P_3, 1 on Fp_1 and 3 on Fp_2) across 13 sessions. The average type II RPs' onset was estimated to be -377.3 ms (SD = 142.4) using the MN method and -421.1 ms (SD = 223.4) using the $RP_{90\%}$. Figure 16b demonstrates typical type II RP.

The type III RPs (negative potentials with a late onset) were found to be almost as common as type I RPs. We found **19 type III RPs** in 11 series (6 on C_z, 4 on C_3, 3 on C_4, 2 on P_3, 2 on Fp_1 and 2 on Fp_2) across 10 sessions. The average onset was -250.7 ms (SD = 84.2) using the MN method and -249.7 ms (SD = 201.9) using the $RP_{90\%}$ method. We found that the type III RPs can be further divided into low amplitude RP III (RP with late onset and low voltage on its peak) and high amplitude RP III (RP with late onset, sharp shape and high voltage on its peak). Figure 16c demonstrates typical low amplitude type III RP, Figure 16d depicts typical high amplitude type III RP.

Figure 16a
Type I RP, C_z, participant 5, 39 valid trials
(MN onset -679 ms, $RP_{90\%}$ onset -484 ms)
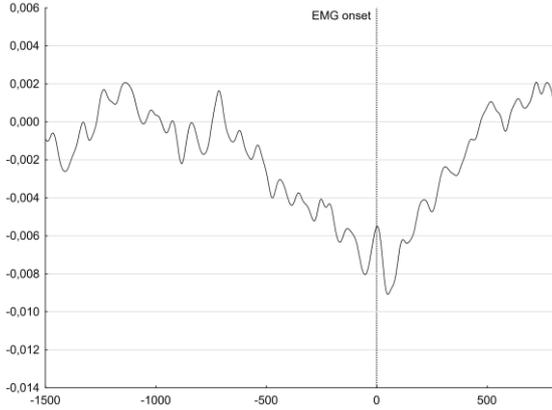
Figure 16b
Type II RP, C_z, participant 5, 36 valid trials
(MN onset -349 ms, $RP_{90\%}$ onset -343 ms)
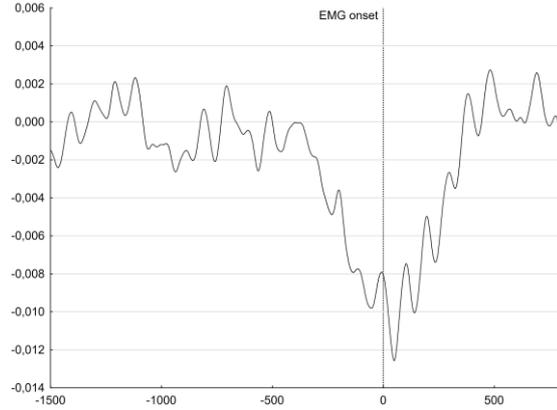
Figure 16c
Low amplitude type III RP, C_z, participant 1,
21 valid trials
(MN onset -237 ms, $RP_{90\%}$ onset -152 ms)
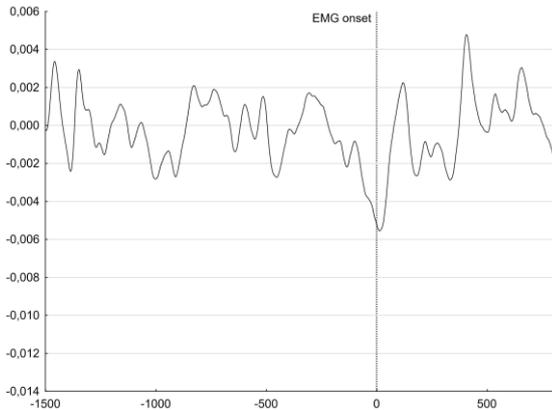
Figure 16d
High amplitude type III RP, C_z, participant 8,
40 valid trials
(MN onset -279 ms, $RP_{90\%}$ onset -232 ms)
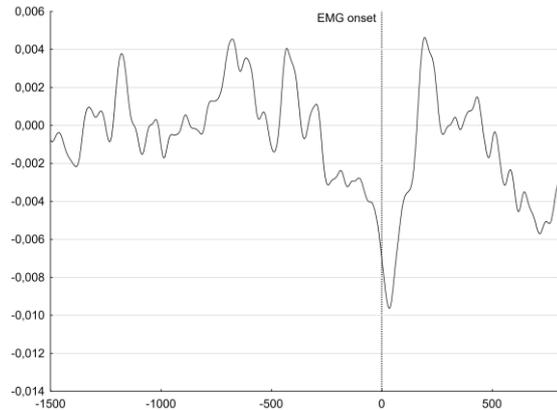
Figure 16: Examples of type I, type II and type III RPs on the C_z electrode. The plots depict averaged trials in one series.

We investigated the differences between the MN and $RP_{90\%}$ methods of RP estimation in the M and W series. On average, the MN method estimated earlier RP onsets by mere 3.0 ms (SD = 224.1). The correlation between the methods' estimates is moderate, r = 0.50. The $RP_{90\%}$ method recognized every RP identified in the MN estimation. The number of the opposite cases is, however, substantial—of the 226 plots in which no RP was found using the MN method, the $RP_{90\%}$ identified 130 RPs (57.5 %). See Figure 17a and 17b for examples of these $RP_{90\%}$ "false positive" results.[10]

---

[10] In this context, we call the result "false positive" if the $RP_{90\%}$ detects an RP where RP_{MN} does not.

Figure 17a
"false positive", C_3, participant 6, 30 valid trials
($RP_{90\%}$ onset -888 ms)
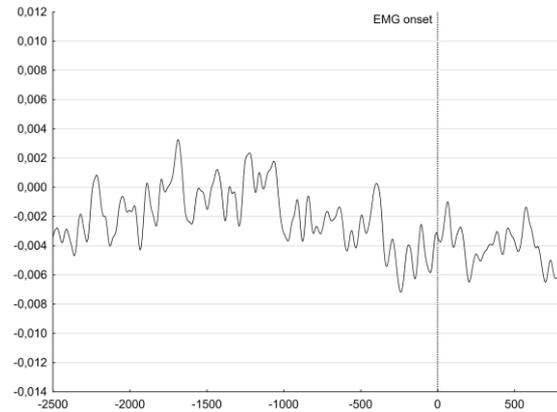
Figure 17b
"false positive", C_z, participant 6, 38 valid trials
($RP_{90\%}$ onset -420 ms)
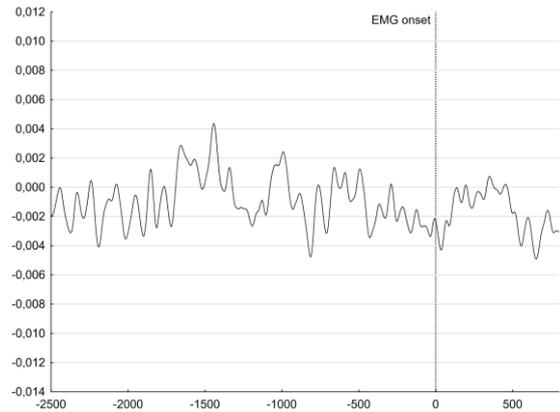
Figure 17: Examples of "false positive" results using the $RP_{90\%}$ method. In this context, I consider the eye-ball detection to be the "true positive".

ERPs in the P and Pv series

In both the P and Pv series, a specific type of RP with extremely early onset was expected. Out of the 75 valid EEG plots from the **P series** we found **2 type I RPs** (in 2 series), **14 type II RPs** (in 5 series), **1 type III RP** (low amplitude) and **25 RPs identified as the pre-set RPs** (found in 9 series; the pre-set RPs most closely resembled the irregular RP shape depicted in Figure 15c). Therefore, type II and pre-set RPs are the predominant RP types present in the P series. The type II RPs' onset was on average -406.8 ms (SD = 127.0) using the MN method and -376.4 ms (SD = 166.3) using the $RP_{90\%}$ method. These type II RPs did not exhibit any notable differences from the type II RPs acquired in the M and W series (for an example, see Figure 18a, compare to Figure 16b). The pre-set RPs' onset was on average -1267.4 ms (SD = 185.9) using the MN method and -1035.6 ms (SD = 144.5) using the $RP_{90\%}$ method (one of the 25 pre-set RPs identified using the MN method was undetected by the $RP_{90\%}$). A striking characteristic of the pre-set RPs in the P series was that they tended to reach the peak negativity long before the target time (see Figure 18b). Another interesting fact is that not a single pre-set RP was ever identified on the Fp_1 and Fp_2 electrodes (in neither of the P and Pv series); in the P series, 7 pre-set RPs were detected on C_z, 6 on C_3, 4 on C_4 and 8 on P_3.

Figure 18a
Type II RP in P series, C_z, participant 2, 39 valid trials
(MN onset -663 ms, $RP_{90\%}$ onset -365)
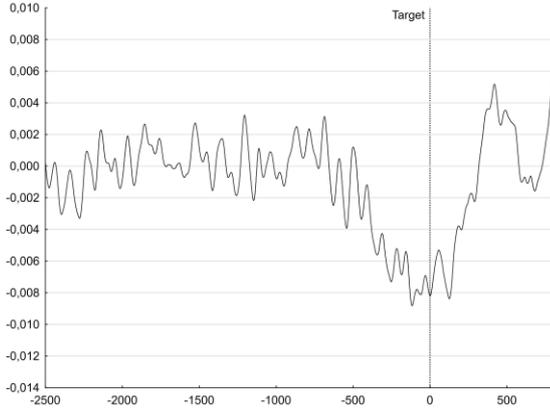
Figure 18b
Pre-set RP in P series, C_z, participant 7, 39 valid trials
(MN onset -1177 ms, $RP_{90\%}$ onset -956 ms)
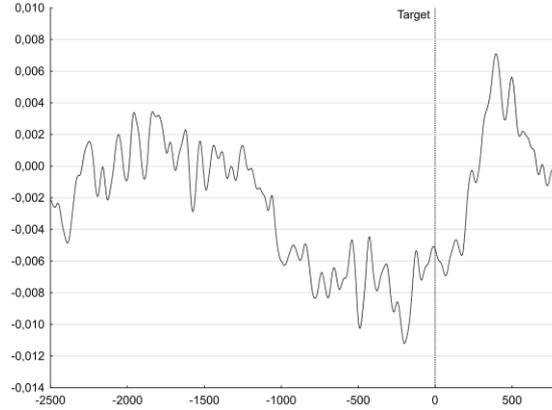
Figure 18c
Pre-set RP in Pv series, C_3, participant 7, 14 valid trials
(MN onset -801 ms, $RP_{90\%}$ onset -1073)
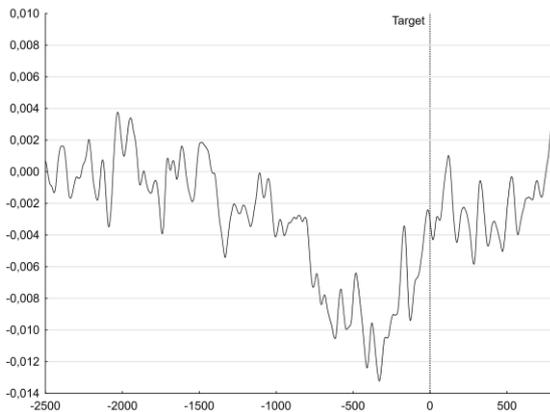
Figure 18d
Pre-set RP in Pv series, C_z, participant 3, 18 valid trials
(MN onset -1307 ms, $RP_{90\%}$ onset -1081 ms)
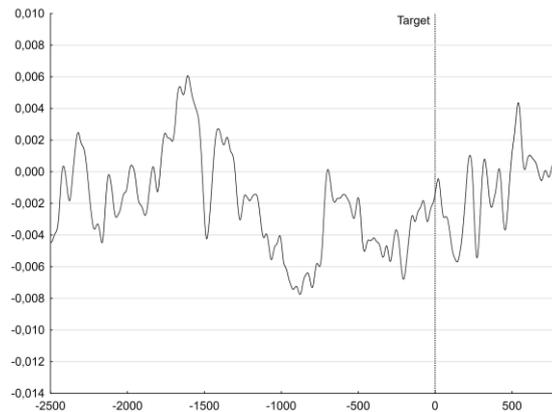
Figure 18: Examples of different RP types identified in the P and Pv series.

In the **Pv series**, we identified **2 type III RPs** and **18 pre-set RPs** (in 7 series) out of 75 valid EEG plots. The mean pre-set RP onset in the Pv series was -1192.1 ms (SD = 180.2) using the MN method and -989.6 ms (SD = 492.6) using the $RP_{90\%}$ method. There are some notable differences between the pre-set RPs in the P and in the Pv series. First, even though Figure 15d demonstrates a potential similar to type I RP, the averaged plots of most individual Pv series from individual electrodes suggest that the vetoed pre-set RP tended to fully return to baseline before the target time was reached (see Figure 18c), sometime by even more than 500 ms. This might be the reason why the $RP_{90\%}$ method did not detect 2 of the 20 pre-set RPs detected by the MN method. Second, the veto pre-set RPs tended to be preceded by a short positivity in some series; the following RP usually had lower amplitude (see Figure 18d).

ERPs in the S series

The EEG recordings in the S series are not supposed to contain any RP negativities, but P300 positivity following the stimulus delivery is expected. Nevertheless, the examiners detected 8 RPs across 114 valid EEG plots in the S series (3 type II RPs, 2 type III RPs, 3 pre-set RPs).

The **P300 wave was detected in 59 of the 114 plots**, in 18 separate series. The four examiners analysing the P300 mostly agreed on the P300 wave detection—30 waves were detected by all four examiners, 14 waves were detected by three examiners, 7 by two and 7 by only one examiner (the waves detected by two and fewer examiners must have been confirmed with special care by the researcher who merged the assessments into a single conclusion, i.e. myself). The P300 waves in the S series were mostly found on the C_3 and C_4 electrodes (15 occurrences on C_3, 16 on C_4); 13 waves were found on P_3, 8 on C_z, 3 on Fp_1 and 4 on Fp_2. For an example of a detected P300 waveform, see Figure 19a.

Figure 19a
P300 in S series, C_3, participant 3, 41 valid trials

Figure 19b
P300 and RP in P series, C_z, participant 8, 40 valid trials
(MN onset -582 ms, $RP_{90\%}$ onset -1169 ms)
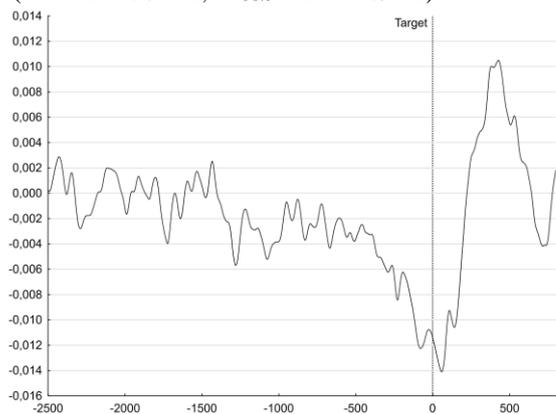
Figure 19: Examples of the P300 waves.

The examiners also found 16 P300 waveforms in other than the S series, namely in 10 plots within five W series and in 6 plots within three P series. These P300 waveforms might have in some cases been misjudged rebounds of preceding RPs, because 10 of the 16 waves were detected together with an RP (for an example, see Figure 19b).

Participant	Session	Mode of recall	Introspective reports (ms)									RP onsets on Cz (ms)					
			W			M			S			RP in W series			RP in M series		
			M	SD	n	M	SD	n	M	SD	n	Type	RP_{MN}	$RP_{90\%}$	Type	RP_{MN}	$RP_{90\%}$
1	2	A	-104	179	38	-79	127	40	+3	148	40		no RP	-33		no RP	-2500
	3	O	-19		41	-14		41	+5		41		no RP	no RP		no RP	no RP
	4	A	-151	89	37	+8	100	38	+48	106	40	II	-216	-742	II	-346	-270
	5	O	-49		41	+47		41	+43		41		no RP	no RP		no RP	-14
	6	A	-342	140	26								no RP	-470			
	7	O	-131		41							III	-237	-152			
2	2	O	-205		41	-70		41	-3		41	I	-735	-589		no RP	-118
	3	A	-512	287	34	+85	92	29	+127	118	40	I	-517	-418		no RP	-888
	4	O	-256		41	-58		41	+43		41	not enough valid trials				no RP	-112
	5	A	-277	246	21	+32	58	27	+148	102	40	II	-272	-409	II	-369	-324
	6	O	-183		41							II	-433	-334			
	7	A	-233	190	25								no RP	-795			
3	2	A	+32	83	25	+44	73	25	+134	113	40		no RP	-401		no RP	no RP
	3	O	+73		41	+54		41	+28		41	low EEG quality			low EEG quality		
	4	A	+43	86	34	+73	56	33	+131	107	40	low EEG quality			low EEG quality		
	5	O	+80		41	+71		41	+58		41	III	-393	-331	II	-309	-477
	6	A	+79	70	23								no RP	-354			
	7	O	-3		41								no RP	-803			
4	2	O	-41		41	+47		41	+58		41	low EEG quality			low EEG quality		
	3	A	+13	116	38	+173	135	38	+121	121	40	low EEG quality			low EEG quality		
	4	O	+70		41	+69		41	+58		41	low EEG quality			low EEG quality		
	5	A	+16	126	33				+120	90	40	low EEG quality			low EEG quality		
	6	O	+54		41							low EEG quality					
	7	A	+76	110	30								no RP	-2			
5	2	A	+11	234	39	+126	224	31	+236	156	40	II	-349	-343	I	-582	-937
	3	O	-46		41	-3		41	+50		41	I	-789	-444	I	-504	-859
	4	A	+59	119	39	+95	161	39	+212	162	40	II	-391	-270	I	-679	-484
	5	O	+4		41	+5		41	+88		41	I	-858	-718	II	-420	-337
	6	A	-2	122	32							low EEG quality					
	7	O	-37		41							low EEG quality					
6	2	O	-71		41	-93		41	-93		41	low EEG quality			low EEG quality		
	3	A	-44	99	34	+71	61	37	+230	118	40	low EEG quality			low EEG quality		
	4	O	-246		41	-147		41	-153		41	low EEG quality			low EEG quality		
	5	A	-257	129	30	+14	68	35	+272	212	40		no RP	no RP		no RP	-360
	6	O	-127		41								no RP	-71			
	7	A	0	84	37								no RP	-420			
7	2	A	-195	143	39	-94	89	40	+117	189	40	low EEG quality			low EEG quality		
	3	O	-59		41	+21		41	+88		41	low EEG quality			low EEG quality		
	4	A	-80	78	39	+54	104	40	+263	118	40	low EEG quality			low EEG quality		
	5	O	-93		41	+126		41	+110		41	II	-352	-279		no RP	-305
	6	A	-34	94	35							I	-1037	-864			
	7	O	+90		41								no RP	-510			
8	2	O	-124		41	+98		41	+103		41		no RP	-34	III	-270	-27
	3	A	-149	93	40	-52	83	39	+99	81	40	III	-279	-232	III	-147	-90
	4	O	-195		41	-152		41	+65		41		no RP	no RP		no RP	-3
	5	A	-164	74	40	-85	79	40	+74	88	40		no RP	no RP		no RP	-22
	6	O	-172		41							III	-334	-194			
	7	A	-212	100	40							I	-563	-451			

Table 9: Overall summary of the data obtained in the W, M and S series. Session 1 is not displayed, as it served training purposes only. The M and W reports are calculated in relation to the EMG onset, the S reports to the skin stimulus delivery; their SDs are calculated for the absolute mode of recall only; in the order mode of recall, only mean estimates are presented. In the RP onsets section, both RP_{MN} and $RP_{90\%}$ on C_z are presented, but the RP type is stated only when the RP was detected using the eye-ball MN method. Label "low EEG quality" signifies that C_z electrode collected excessive noise, so the signal could not be analysed; "no RP" means that the RP was not detected; "not enough valid trials" means that the averaged EEG plot was calculated from fewer than 10 trials.

5.3.4 Interindividual differences

It has been suggested in chapter 5.3.2 that the participants differed from each other in their introspective reports. In this chapter, I would like to summarize most of the available data into a comprehensive Table 9, which can be compared to Libet's results table (see Libet, Gleason, et al., 1983, p. 630, Table 1).

Table 9 shows some important points, which I discuss in chapter 5.4.4. First, some of the mean W reports referred to time after the EMG onset, especially in the cases of participants 3 and 4. Second, the introspective reports generally differed substantially between the participants, as was already suggested in chapter 5.3.2. Third, the noisiness of the EEG recordings led to a rejection of many RP measurements. Furthermore, many of the remaining RP plots contained no RP detected by the MN method and in some cases neither by the MN and $RP_{90\%}$ method.

5.3.5 Series-specific effects

In this chapter, I report how the participants answered our control questions and what possible effects could potential exceptionalities have on other results. The two questions are stated in chapter 5.2.5 and always pertained M and W series. **The first question** asked whether the participant felt that **the movement was pre-planned** and not spontaneous. In most sessions, the participants stated that they did not feel any pre-planning before moving the finger. If the participants answered that they did (which happened in 17 of the 48 sessions), they usually stated that the pre-planning occurred in 1 to 4 trials. In two cases, the reported frequencies of the pre-planning were higher than 4. In the first case, participant 8 reported that approximately 8 trials in one W(A) series were pre-planned—the ERP detected in this series on C_3, C_4 and P_3 was type II RP with MN onsets -699 ms, -699 ms and -898 ms, respectively; the W(A) reports in the series were earlier than the sample average (M = -212.4 ms, SD = 100.3). In the second case, the same participant reported approximately 6 trials in one M(A) series to be pre-planned—this time, the RP detected on C_z, C_3, C_4 and P_3 was type II with MN onsets -309 ms, -315 ms, -302 ms and -298 ms, respectively; the M(A) reports in this series were notably earlier than the sample average (M = -85.0 ms, SD = 78.9).

The second question asked whether the participant **was surprised by his or her own movement**. In 30 of the 48 sessions, the participants stated that they were surprised by the movement in at least one trial. Reported frequencies of this happening were overall similar

to the answers to the first question—participants usually reported approximately four movements in a series to be surprising. Participant 1 reported six times that the situation happened in 10 and more trials in a single series. We examined the results from these six series (3 W(A), 1 M(A), 1 W(O) and 1 M(O) series). We inspected 36 corresponding EEG plots and found that the RPs were rarely detected using the MN method. If any RPs were found (in 9 cases), they were predominantly type II and type III with low amplitude and late onset (with an exception of type I RPs found on the Fp_1 and Fp_2 electrodes in one of the W(A) series). The introspective reports in said series were as follows: mean M(A) report was 8.1 ms (SD = 99.9), mean W(A) report was -182 ms (SD = 169.9), M(O) mean estimate was 46.9 ms and W(O) mean estimate was -48.9 ms; these results do not seem extraordinary in any way among other participants' introspective reports.

5.4 Discussion

This study aims to replicate the outcomes of Libet et al. (1982), Libet, Gleason, et al. (1983), Libet, Wright, et al. (1983) and Libet (1985) by conducting a study designed to follow the original methodology as closely as possible. The experiment is complex and as our data suggest, its outcomes are substantially dependent on data collection and analysis procedures. In the following six chapters, I discuss technical issues and effects of our methodological choices, as well as the implications of our replication and other recent empirical studies for the validity of Libet's experiment.

5.4.1 Procedure and instruction discussions

Our replication differs from Libet's methodology in several ways. One of them is the sample—we tested eight participants instead of six (Libet et al., 1982, p. 323; Libet, Gleason, et al., 1983, p. 624); this may seem as an irrelevant change, but for future Libet-style experiments analysing both the M and W series in both the A and O modes of recall, we strongly recommend using sample sizes divisible by 4 to allow for complete experimental conditions rotation (see chapter 5.2.5, Table 5). I would also recommend conducting a full-scale replication with a much larger sample and attempt to find the factors responsible for the large interindividual variability. The reason is that our results and results of other researchers, Libet including, suggest that some participants tend to differ considerably in their introspective reports from the rest of the sample (e.g. later W reports of participants 3 and 4 in our study, see Table 9; later M and S reports of participant B.D. in Libet, Gleason,

et al., 1983; later W reports of participant IB in Keller & Heckhausen, 1990; earlier W reports of participant 5 in Verbaarschot et al., 2015). Simple grand average results from the introspective data are therefore meaningless, because they do not account for the interindividual differences (that is also why we used the mixed-effect linear model to analyse the W, M and S reports, see chapter 5.3.2).

Another problematic element of the experimental procedure is the instruction, especially when asking for the introspective report, which I signify as W in this text. In Libet's original experiment, the participant was instructed to "*report the time of appearance of his conscious awareness of 'wanting' to perform a given self-initiated movement*" (Libet, Gleason, et al., 1983, p. 627), but Libet also states that the participants described the experience as an "urge", "intention" or "decision". Soon et al. (2008, p. 543) instructed the participants to "*press a button as soon as they felt the urge to do so*", but to report "*when their motor decision was consciously made*". Trevena & Miller (2002, p. 172) asked the participants to report the dot position "*at the time of the decision to 'go now'*". Verbaarschot et al. (2015, p. 301) asked the participants "*to report the onset of their intention to act*". Caspar and Cleeremans (2015, p. 4) instructed the participants "*to report the location that the black spot occupied at the time they had first decided to press the key*". Most of these studies diverged notably in the acquired W results which suggests that the instruction formulation might influence the results significantly. In our case the instruction was to report the clock time when the participant "*realized the first urge to press the mouse button*".[11] We decided to use the word "urge" for following reasons. First, Libet, Gleason, et al. (1983, p. 627) stated that "*subjects usually settled for the words 'wanting' or 'urge'*"; we intended to use one of these and—as "wanting" does not sound natural in Czech sentence—we decided to choose the "urge" option. Second, Pockett & Purdy (2011) argue that there is a difference in the experiment results when instructions contain either "decision" or "urge" and that the "urge" instruction replicates Libet's results.

Another potentially significant deviation from Libet's methodology is that we did not provide any feedback to the participants based on their S reports. Our reasons were explained at the end of chapter 5.2.4. As Gomes (1998, p. 590) suggests, providing feedback on how accurate the participant is in every S series throughout the experiment will certainly change

---

[11] "(…) v okamžiku, kde jste si uvědomil/uvědomila první nutkání stisknout tlačítko" in Czech.

the results leading to large inconsistencies in data obtained in the first few sessions compared to data acquired in the last sessions. Furthermore, based on the results of Danquah et al. (2008), we can assume that the feedback training would have varying effects on the M and W reports based on the stimulus modality in the S series. Based on these obvious problems, we decided not to tell the participants how accurate they were in the S series, nor to perform any S trainings at the beginning of the sessions (as these would naturally require us to provide the feedback to have any training effect). One could argue that the fact that we did not provide the feedback was the cause for the strikingly late S reports in the S(A) series. However strange it may appear that this would influence only the S(A) reports and not the S(O) reports, I admit that the S series might indeed require training to be valid. Based on this intricacy, we also refrained from correcting the M and W reports using the S reports, as Libet suggested (Libet, Gleason, et al., 1983, p. 631).

5.4.2 Technical equipment discussions

Our technical equipment allowed us to perform some of the experimental procedures more precisely. The display used was a normal LCD computer display, which compared to the cathode-ray oscilloscope (CRO) used in Libet's case (Libet, Gleason, et al., 1983, p. 625) offers better contrast (for that reason we also used black and white colours instead of the grey and bright green displayed by the CRO, see Figure 2 in chapter 2.2.3). Also, we presumably increased the reporting precision in the A mode of recall by allowing our participants to click on the reported location.

We did not make changes to the EEG measurements except replacing the C_c and C_i electrodes with C_3 and C_4 (C_c and C_i locations cannot be measured using the standardized 10-20 EEG cap, see chapter 5.2.3). As it turned out, the eye fixation was not as perfect as Libet, Gleason, et al. (1983, p. 323) asserted—after the EEG analyses, we acknowledged that the EOG measurements would in fact reduce the number of epochs rejected due to the eye-movements. I recommend future researchers to use the EOG to remove the eye-related artefacts from the EEG recordings.

The S series required a skin stimulator, which in Libet's case was electrical (Libet et al., 1982, pp. 323–324). We decided to use a tactile stimulator to eliminate the risk that an electrical impulse would interfere with the EEG data. It might be objected that the tactile skin stimulator is not as precise as the electrical stimulator, because it needs to mechanically move a bolt touching the participant's skin to deliver the stimulus. I agree that this might be

an issue, but we did not find any viable measure to check the stimulator's precision other than our own subjective test. However, to eliminate additional potential error in the timing of the trigger signal (sent through a 3.5 mm audio cable) driving the stimulator, we used a check loop recording the difference in signal input, output and repeated input on the MP150 unit which did not show any latencies whatsoever.

5.4.3 EMG measurements discussions

For technical reasons addressed in chapter 5.2.4, the movement in our study had to be in the form of a mouse click. As argued earlier, this may allow us to investigate some additional issues. Button press is used in many Libet-style experiments and results of this type of response are expected to be equivalent to Libet's original flexion of the wrist or fingers (Libet, Gleason, et al., 1983, p. 625). As it turned out, the mouse click might have a few issues. First, the click produces an auditory feedback, which might shift the introspective reports, as showed by Banks and Isham (2009), so it requires additional measures (such as the earplugs used by us) to overcome. Second, even if the auditory feedback is eliminated, the mouse still provides a slight haptic feedback when pressed—this might raise the question whether the participants relate the M and W reports to the EMG onset or the mouse click. Third, because the movement is small, it happened multiple times in our experiment that the EMG onset was not registered by the electromyograph if the participant conducted the movement not rapidly enough or if the electrodes were not attached precisely (which would be a smaller problem if the movement was a flexion of the whole wrist, because the number of activated muscle fibres in such movement is significantly higher). Therefore, our conjecture stated in chapter 5.2.4 that the mouse click provides a better-bounded movement more suitable for Libet's task seems to be incorrect.

The movement issues are also related to our results pertaining the EMG onsets. Besides the 286 trials containing no detectable EMG onset in the M, W and P series, another 318 trials contained EMG preceding the mouse click by more than 150 ms or following it by any amount of time. These results should be further discussed. The EMG onsets might have been undetected for two main reasons: either the integrated EMG level did not reach sufficient threshold to be detected or early artefacts were present in the recording obscuring the EMG onset. The early EMG onsets (< -150 ms) were most probably caused by early supraliminal muscle twitches, while the late EMG onsets (> 0 ms) were found presumably due to insufficiently abrupt movement initiation (causing the baseline to rise gradually, never

exceeding the threshold based on its IQR, see chapter 5.2.6). Even though these errors were not common among the 3879 M, W and P trials, they may cause issues when averaging the EEG, which relies on the temporal alignment based on the EMG onsets (their non-detection reduces the number of valid epochs entering the averaging procedure). Seemingly, this problem could have also been worked around if the wrist flexion was used instead of the mouse click; nevertheless, the mouse click also serves as a kind of a control mechanism checking that an early artefact is not detected as the EMG onset. Each of the technical solutions seems to have its advantages and disadvantages.

As stated in chapter 5.3.1, we also compared the EMG onset data related to the mouse click to the results of Haggard and Eimer (1999). Our results suggest that absolute majority of the EMG onsets precedes the mouse click by 30–90 ms, but that some outliers are present on both ends of the range spanning from -330 to +106 ms. It may be needed to perform a study dedicated solely to this issue.

For all the aforementioned reasons, **I recommend caution when using the mouse click to time the introspective reports**.

5.4.4 Introspective reports discussions

As stated in chapter 5.4.1, I have doubts that presenting grand averaged introspective reports is meaningful. However, to make our results comparable to those of Libet, we presented raw mean reports (calculated from the series means) for M(A) (M = 30.9 ms, SEM = 79.8), M(O) (M = 0.0, SEM = 83.5), W(A) (M = -101.2, SEM = 151.1), W(O) (M = -70.3, SEM = 104.7), S(A) (M = 146.0, SEM = 77.3) and S(O) (M = 34.1, SEM = 69.4).

Libet's mean M report was -86 ms (Libet, Gleason, et al., 1983, p. 631), which is by 117 ms earlier than our M(A) report and by 86 ms earlier than our M(O) report. There are other studies reporting their mean M time, with a large amount of inter-study variability: -89 ms (SD = 118, Haggard & Eimer, 1999), 19.8 ms (SD = 39.0, Sirigu et al., 2004, healthy participants), 5.2 ms (SEM = 18.6, Pirio Richardson et al., 2006, healthy participants), -91 ms (SD = 92, Moretto, Schwingenschuh, Katschnig, Bhatia, & Haggard, 2011), -59.9 ms (SE = 5.3, Caspar & Cleeremans, 2015). Compared to all the presented results, our data seem to be generally shifted towards later values, but the difference is not significant in many cases. We can safely rule out that the difference is due to omission of the S series trainings, because no other studies than the one conducted by Libet included it. It seems that the shift may stem from the instruction differences or methodological measures

(such as making the report by clicking into the clock face). However, I consider our results to be more valid than those of Libet, because the M reports of our participants are on average far more precise than those of Libet's subjects.

Libet's W reports were -204 ms on average, which is by 105 ms earlier than our mean W(A) report and by 134 ms earlier than our mean W(O) report. The W reports in our replication seemingly differ from Libet's original results, but 95% confidence interval, derived from our mixed-effect model presented in chapter 5.3.2, does include Libet's -204 ms, as well as results found by other researchers (e.g. Trevena & Miller, 2002; Pirio Richardson et al., 2006, healthy participants; Schurger et al., 2012; Verbaarschot et al., 2015; Caspar & Cleeremans, 2015). On the other hand, there are also studies with mean W reports earlier than the lower bound of our CI (e.g. Keller & Heckhausen, 1990; Haggard & Eimer, 1999; Sirigu et al., 2004, healthy participants; Edwards et al., 2011, healthy participants; Moretto et al., 2011, healthy participants). It seems that the W reports are extremely variable. I suggest that these discrepancies are to be expected due to the fact that there is large interindividual variability in the W reports (see chapter 5.3.2) and that many studies use small samples. The question is why the reports of our participants differ from Libet's original findings. The reasons may be that we did not train the participants' precision in the S series (which, as stated in the previous paragraph, seems unlikely) or that we provided different instructions than Libet, as Libet did not publish the whole instruction text; neither of these, however, necessarily means that our W reports are invalid—it may simply mean that our participants understood the W concept differently than Libet's participants or the participants in other studies. Additionally, I also have to address one issue already mentioned several times—Trevena & Miller (2002) reported that about 40% of the W reports were found to be later than the movement onset. In our study, we made a similar finding—249 of all 808 valid W reports (30.8%) were found to follow the EMG onset, while the mean W reports of participants 3 and 4 followed the EMG onset almost exclusively (see chapter 5.3.4, Table 9). In the discussion of Study 1 (chapter 4.4) I have offered a possible explanation for the late W reports in context of our finding that if no M reports are present in the study design, then the participants might tend to provide late W reports because they confuse the impression of (presumed) urge and the movement itself. However, in this Study 2, the M reports were present and thoroughly explained, yet some participants still

provided notably late W reports. This suggests that the omission of the M reports is not the only possible explanation for the late W reports.

In accord with Study 1, we found that the W reports obtained in sessions in which the W task followed the M task were earlier than the W reports obtained before the M task in the respective session; however, the difference was not statistically significant. Our conclusion compliments Libet's results (Libet, Gleason, et al., 1983, p. 632) which also showed the same trend. The most probable reason why the results do not replicate the order effect found in Study 1 is that in this case, the participants were already familiar with both the M and W tasks when making the W reports which were analysed.

Libet's grand averaged S report was -47 ms. This is by 193 ms earlier than our mean S(A) reports and by 81 ms earlier than our mean S(O) report. In this case, I am prone to believe that the difference is indeed caused by omission of the S precision training. However, if this is the case, the training seems to influence the S(A) reports far more than the S(O) reports. In fact, I believe that the O mode of recall might generally be more precise, given the fact that the (O) reports have universally lower SEM and that the mean S(O) and M(O) reports are notably closer to the target event (stimulus or the EMG onset) than the S(A) and M(A) counterparts. This also rules out the possibility that the strikingly late S(A) reports are caused by the unknown stimulator latency (see chapter 5.4.1), because if this was the case, then the S(O) reports should have also been affected (there is no reason to assume that the stimulator would exhibit different latencies in the S(A) trials than in the S(O) trials). Nevertheless, the skin stimulus is delivered by moving the bolt towards the participant's skin and then retracting it—we can speculate that the S(A) reports could somehow lead the participant to assess the time of the "end of the touch", while the S(O) task would encourage the participant to report the time of the "beginning of the touch". This may require further testing using various skin stimulators.

We also found significant differences between the M(A), W(A) and S(A) reports. I find it important that the W(A) reports precede the M(A) reports significantly, suggesting that our participants did not usually confuse the W impression for the M impression. However, it should be noted that the M and W reports of participant 3 did not significantly differ, complimenting the idea presented in Study 1 suggesting that the W experience might have, for some individuals, the same meaning as the M experience.

By the end of this section, I feel obliged to point out one specific problem, which seems to significantly encumber the "libetian" discussion—the SEMs can be assessed using two different calculation methods which may lead to vastly different results. If the SEM is calculated as the standard deviation of the series means, then the procedure corresponds to Libet's original procedure. On the other hand, if the SEM is calculated using the formula:

$$SEM = \frac{SD}{\sqrt{n}}$$

then the procedure assumes that there is no difference between the series means (which is not true, as suggested in Table 9) and may lead to an underestimated result. I therefore recommend calculating the SEM using the first procedure when reporting the introspective data in Libet-style experiments.

5.4.5 ERP discussions

When analysing the EEG data, we found several loopholes in Libet's original process. Many of the common RP analysis problems are discussed elsewhere (e.g. Verbaarschot et al., 2015). I will therefore address only the problems which seem to be relevant to Libet's suggestions (Libet et al., 1982; Libet, Gleason, et al., 1983).

First and foremost, in many cases we did not find any RPs in the EEG plots where they were actually supposed to be found. Our experience can be summarized by a quotation of Pockett & Purdy (2011, pp. 4–5, bold added): "*When one first begins to investigate the event-related potentials arising in the 2 s prior to voluntary movements, it rapidly becomes clear that **not all experimental subjects generate RPs**. As with many negative findings, the idea of trying to publish this result is soon overtaken by the realization that it would be far too easily rejected on the grounds that everyone can record RPs, so there must have been some technical inadequacy in the recording sessions where none was seen.*" I am indeed aware of some technical limitations to our EEG recording equipment (see chapter 5.5), but I would like to support Pockett and Purdy in their opinion and to suggest that in many circumstances, the averaged EEG plots might, indeed, not exhibit any RP detectable by an eye-ball inspection when it is in fact expected. When inspecting the EEG plots using the MN method (i.e. eye-ball inspection), we agreed to assess an RP only if it satisfyingly fitted the description in our "Guide to the RP identification" (see Appendix 6). Our reason for this was to be sure that we analyse an RP and not some non-specific negativities present even in a resting EEG (in fact, we found some RPs preceding the skin stimulation in the S series which

makes it clear that the RP may be falsely identified where it is not supposed to be identified, even while doing it with a great caution). We used independent assessments made by five researchers, merged the assessments and then let another examiner check the results. However, the results might have been influenced by the fact that when merging, some assessments had to be corrected because they violated an important rule in the "Guide" or because the examiners did not agree on the onset timing. In both cases, I, as the merging examiner, had to subjectively choose one of the assessments, which admittedly introduced some amount of the previous subjectivity into the process.

Another potential factor reducing the number of identified RPs was relatively high high-pass filter (0.5 Hz), which was, however, found to be necessary to remove slow voltage drifts present in the raw recording, but could also move the onset notably (see Verbaarschot et al., 2015, Figure 6). One could also wonder why we made the MN assessments with millisecond precision and not by 100 ms step as in the Libet's case (see Libet, Gleason, et al., 1983, p. 630, Table I); the reason was that RPs sometimes had sharp onsets, so we found it useful to time the onsets precisely (which we did using a cursor which displayed the timestamp of a specific data point). To summarize the discussion of MN method, we are aware that some negativities might have in fact been RPs, but as it was often difficult to recognize an RP in an ambiguous plot, we chose to use a rather cautious approach.

I argue that the RP$_{90\%}$ method seems to be also problematic, due to multiple factors. First, it seems to detect the RP onset systematically later than it is supposed to be detected using the MN method (see Verbaarschot et al., 2015, Figure 4A and 4B). Second, its result is heavily dependent on the baseline, which needs to be set prior to the calculation; we had to use different baselines for the M, W and S series than for the P and Pv series, because too early-set baseline in the M, W and S recordings distorted the onset estimates both ways due to early voltage shifts while the late baseline in the P and Pv series was calculated from a segment in which the RP negativity was already rising. We found that the RP$_{90\%}$ was more sensitive to negativities which were not identified as RPs in the MN method, but I find it open to discussion which of the two methods represent the reality better.

Even though the RPs were not found by the MN method in all relevant series, our recordings were indisputably valid, as argued in chapter 5.3.3 and demonstrated in grand-averaged plots in Figure 15. Most strikingly, there seems to be a pronounced RP negativity in the M/W, P and Pv series while no such negativity is present in the grand-averaged EEG

plot from the S series. It is also interesting to confirm Libet's assertion than an RP precedes a target time in the Pv series, even though no supraliminal movement was made. The grand-averaged RP in the Pv series (see Figure 15d) corresponds strikingly to Libet's description of the vetoed RP: "*In these series a ramplike [sic] pre-event potential was still recorded during >1 sec before the preset time (...), even though no actual muscle activation occurred (...). This resembles the RP of self-initiated acts when preplanning is present*" (Libet, 1985, p. 538).

Before we analysed the individual EEG plots, we defined the four RP types (see chapter 5.2.6). However, we encountered a problem that Libet characterized the types by both their shape and their onset (Libet et al., 1982, p. 326), which led to many paradoxical findings of inconsistent shape and onset (e.g. a dome-shaped type II RP with an onset -300 ms, which is too late to be a type II RP). These inconsistencies were solved by prioritizing the shape to the onset, with an exception of type III RPs, which we defined as any late-onset negativities that cannot be regarded type II RPs. This might have led to a rather large number of type III RPs found in our data compared to their scarcity in Libet's findings (Libet, Gleason, et al., 1983, p. 630, Table I). Overall, I understand the benefit of classifying the RPs, but we did not find a clear connection between type I RP and the pre-planning impression or between type III RP and the impression of surprise by own movement (as suggested by Libet et al., 1982, pp. 327–330), even though it seems suggestive that the „surprise" impression may be connected to a late RP onset and low amplitude (see chapter 5.3.5). Our method of asking the participants whether they felt any pre-planning or spontaneity was rather crude; it might be advisable to use real-time inquiring as proposed by Verbaarschot et al. (2015, p. 304).

Our findings show that if an RP was found, it almost universally tends to precede the W report. However, Table 9 also shows that in some series of one participant, the mean W report was remarkably close to the RP_{MN} onset (participant 2, sessions 3 and 5). It seems suggestive that the "*RP does not precede the intention to act in all participants*" (Verbaarschot et al., 2015, p. 310).

5.4.6 Libet's experiment interpretations discussions

Because our data differ notably from Libet's original outcomes, it seems advisable to discuss the potential implications for the validity of Libet's experiment. First and foremost, it seems to remain true that the RP generally precedes both the M and W reports acquired by the rotating spot method. Libet interprets the RP as a neural precursor of a decision to move

(Libet et al., 1982; Libet, Gleason, et al., 1983). However, our outcomes do not rule out some alternative interpretations of the RP. For instance, Alexander et al. (2016) suggest that a negative potential remarkably similar to the RP can also be observed in a case of purely cognitive decision, which lacks any detectable muscle movement. Schurger et al. (2012) offered an interpretation of the RP as a result of averaging spontaneous fluctuations in EEG occasionally building up to an activation threshold. However, as Schurger et al. point out, this interpretation pertains only RPs present in the task with self-initiated movement and thus cannot explain the RP waveform in the P and Pv series (see my discussion of this fact in chapter 3.4); nevertheless, this does not mean that spontaneous fluctuation cannot at least partially contribute to the RP generation in the self-initiated movement tasks. Because our results do not contradict these alternative explanations, the challenges for the Libet's interpretations implied by them remain valid and should be further investigated.

Another point of interest in this discussion is our confirmation of Libet's suggestion that the RP precedes a vetoed movement (Libet, Wright et al., 1983; Libet, 1985). This finding compliments conclusions by Schultze-Kraft et al. (2016), who also demonstrated that participants are able to veto a movement after the RP onset. In fact, their methodology seems to be even more convincing than Libet's original veto procedure, because the task by Schultze-Kraft et al. allowed the participants to prepare to move at any time and then ordered them to stop the movement when a BCI detected an RP onset (thus avoiding the need to make the movement "pre-set"). Libet (1985) suggests that the conscious veto can actually be considered an instrument allowing us to exert free will, because it can provide a way to block unconsciously arisen urge to act and because it presumably does not have any known neural correlate. However, complete freedom of such veto is in question, as it was demonstrated elsewhere that the conscious veto seems to have a neural correlate as well (Brass & Haggard, 2007). Therefore, even though we managed to replicate Libet's results in the Pv series, we do not have any evidence to support the claim that the conscious veto represents free will (or rather "free won't").

As stated in chapter 5.1, the use of the rotating spot method as means of measuring the introspective reports has been investigated in several papers. While the validity of the rotating spot method for the M reports seems to be convincingly supported (Pockett & Miller, 2007), many studies suggested or demonstrated possible biases in the W reports, i.e. the reports of the urge, intention or wanting to move (e.g. Gomes, 1998; Lau et al., 2007;

Danquah et al., 2008; Banks & Isham, 2009; Study 1 of this thesis). Therefore, it is dubious at least to base any large claims about the free will on the comparison of the W reports and the RP onsets.

5.5 Limitations

Present study contains several technical limitations, some of which were already mentioned in the previous chapters. Here, I will report additional important technical difficulties which may limit our study.

One limitation is a minor flaw in the mean report calculation from the M and W series in the order (O) mode of recall. The flaw is that not all trials in the O mode of recall contained an EMG onset, even though the button was pressed. The estimate might therefore be slightly inaccurate, because we related the reports to the mean EMG onset calculated for the whole series, but I assume that the error was not systematic. This limitation can be overcome in the future research if the EMG, not the mouse click, controlled the Libet's clock software (as in Libet's original experiment).

Other limitations pertain the EEG recordings and analyses. For the recordings, we used the available BIOPAC EEG100C amplifiers designed for spontaneous electroencephalography, not more commonly used acquisition units, such as BioSemi (BioSemi, n.d.). Although we pre-processed the EEG signal as stated in chapter 5.2.6, the EEG contained a large amount of noise and many of the recordings had to be rejected. This might also be the reason why in many of the averaged EEG plots we failed to identify an RP, which could potentially be present but was obscured by the noise. The noise, together with the presence of eye movement artefacts, led to rejecting many single trials from recordings of overall satisfying quality. I would recommend using specialized equipment and combine the EEG recordings with EOG to identify and filter out the artefacts caused by eye movement. I also recommend following additional EEG processing steps, such as re-referencing the signal (see Verbaarschot et al., 2015), many of which were unavailable to us, because we did not record the EEG from the whole scalp.

5.6 Conclusions

Our study's goal was to replicate Libet's experiment and to point out some methodological problems obscured by the experiment's complexity. To my knowledge, a replication as complex as ours was never carried out, and even though we failed to replicate

some elements of Libet's experiment accurately, we found it critical to present our approach and outcomes. The main outcome is that our results from Study 2 replicate Libet's key findings, although substantial differences were found in some of the results' categories, such as the introspective reports or the number of readiness-potentials found. We showed some technical issues in both Libet's methodology and methodology used by later Libet-style experiments. Our data also showed that the results are highly variable and that the discrepancies found between many Libet-style studies are still found even when the original methodology is followed closely.

6 Study 3: Libet's experiment in clinical context

In this chapter I will introduce a currently unpublished study of Libet-style methodology applied in clinical environment. The idea of this research was inspired by several studies suggesting notable differences in introspective reports of participants based on their psychiatric diagnoses.

6.1 Introduction

Despite Libet's experiment being constantly discussed in the fields of cognitive psychology and neuroscience, some findings suggest that its key element—the introspective reports of M and W—might prove useful in the context of clinical diagnostics.

I have already discussed Sirigu et al.'s (2004) research showing that if the Libet's M and W tasks are presented to patients with parietal lesion, the results tend to exhibit unusually late W reports. More specifically, the authors studied three groups of participants: healthy participants (n = 5), patients with selective lesion in the cerebellum (n = 5) and patients with selective lesion in the parietal cortex (n = 5). The participants passed several M, W and S trials in the absolute mode of recall. The movement was objectively recorded using an EMG. EEG was also recorded, but the results are irrelevant to current text. The introspective results showed notable difference between the groups in the W reports. Specifically, the parietal patients reported the W times significantly (reported $p < 0.005$) closer to the movement onset (M = -239.2 ms, SD = 92.9, for healthy participants and M = -314.1 ms, SD = 193.2, for cerebellar patients, but M = -55.0 ms, SD = 132.6, for parietal patients). Interestingly, this effect was not found for the M and S reports. The authors suggested that the parietal lesion causes incorrect internal timing of motor intention, despite general time perception being unimpaired (as shown by the M and S results).

113

Similarly, Pirio Richardson et al. (2006) studied healthy participants (n = 14) and patients with schizophrenia (n = 6) in the M and W tasks in the absolute mode of recall and with reports related to the EMG onsets. Again, the results showed notably (although not significantly, reported p = 0.08) later W reports for schizophrenia patients (M = 10.4 ms, SEM = 75.8) compared to the results of healthy controls (M = -121.6 ms, SEM = 45.6). A notable but non-significant difference was also found for the M reports (M = 5.2 ms, SEM = 18.6, for healthy participants; M = 49.8 ms, SEM = 14.0, for the patients). The authors suggested that the results might be related to impaired sense of agency accompanying most of psychotic disorders.

Edwards et al. (2011) found similar results in patients diagnosed with clinically definite psychogenic tremor (n = 9) when compared to 9 healthy controls. The participants were presented with the M and W tasks in blocks of 30 trials (blocks were preceded by ten training trials for each condition). In this case, the results showed significantly later W reports (p = 0.009), as well as M reports (p = 0.03) in the patient group compared to the controls. The authors accounted for possible motor-related issues in patients with hand tremor but found no significant difference in the reports of movement made by the left and right hand of patients with unilateral tremor. I would cautiously speculate that one possible factor for these unusual results might be the underlying neurotic process, given that the effect was not found to be directly related to the movement disorder.

Moretto et al. (2011) conducted a similar study on patients with the Tourette syndrome. The syndrome is characterised by involuntary tics, which are, however, sometimes described by the patients as being preceded by a feeling of an urge. The authors studied 13 patients diagnosed with the Tourette syndrome and 13 healthy controls. The participants passed two blocks of 30 M and W trials, one condition in each block. The session was preceded by 10 training trials for each condition. The results showed that the patients reported the W time on average 93 ms (SD = 79) before a keypress, whereas healthy participants reported the W time on average 247 ms (SD = 159) before the keypress; the difference was statistically significant (p = 0.005). On the other hand, the difference was not significant for the M reports (p = 0.150). The authors conclude that it seems that the Tourette syndrome patients have delayed experience of intention to move and suggest several possible interpretations for this phenomenon. For example, they suggest that patients suffering from tics disorder might be used to involuntary movements and transfer this impression of non-voluntariness to

stereotyped, but voluntary actions such as pushing a button. Other explanations offered by the authors are that the late W reports might be related to constant effort to supress tics or that getting used to involuntary tics might cause extinction of the learned association between a voluntary intention and an executed action.

These results might overall suggest that there are specific diagnoses, which might be identifiable using the Libet's clock task. If so, then Libet's clock might prove to be a valuable psychodiagnostical tool. However, the studies mentioned above share several downsides. First, they were not originally intended to validate Libet's clock for diagnostical use. Second, the studies employ different methodological procedures and are not strictly standardized; in this regard, it should be noted that results for the control group varied between the studies. Third, the sample size in these studies is rather small and it might be interesting to see similar results obtained on larger clinical samples. Fourth, due to the emphasis on significant results in current publishing practice, it cannot be said that the mentioned diagnoses exhibit said effects and others do not, because it is hard to know whether other diagnoses were tested with non-significant results (and hence failed to be published) or not tested at all[12].

If the phenomenon of later W reports is specific to certain diagnoses, a linking factor should be sought. Some findings were made regarding a relationship between the W reports and personality traits, specifically impulsivity. Caspar and Cleeremans (2015) showed that higher impulsivity seems to lead to later W reports. The authors studied sample of 72 healthy participants and used—besides the Libet's task—several questionnaires measuring impulsivity and other personality traits. Finding of the abnormal W reports in impulsive individuals corresponds with previous findings of abnormal time judgements in interval estimation tasks. Barratt (1983) reported underestimated time intervals in impulsive individuals. Similar conclusion was suggested by trends found in an experiment by Van den Broek, Bradshaw, & Szabadi (1992), in which the participants were asked to reproduce time intervals of various lengths or to discriminate between two intervals. Wittmann, Leland, Churan, & Paulus (2007) linked the phenomenon to the area of addiction, suggesting that stimulant-dependent individuals exhibit high levels of impulsivity as well as altered temporal processing. Admittedly, however, the alteration to the temporal processing found by

---

[12] Admittedly, Pirio Richardson et al. (2006) reported suggestive effects which were not found to be statistically significant, although only in a form of a summary of oral communication.

Wittmann et al. (2007) was most notable in time intervals of 1000 ms and more, which does not completely correspond to the reports of intention to act preceding the action only by several hundreds of milliseconds typical for Libet-style experiments. Nevertheless, it seems suggestive that impulsivity and temporal processing are related.

It should also be noted that if the connection between the W reports and impulsivity is valid, then the W reports might be specifically biased in individuals with addiction (among other diagnoses stated above). That is because impulsivity is generally considered a predisposition factor for alcohol addiction (e.g. see Coskunpinar, Dir, & Cyders, 2013[13]), stimulant addiction (Wittman et al., 2007; Ersche, Turton, Pradhan, Bullmore, & Robbins, 2010), opiates addiction (Nielsen et al., 2012) and behavioural addictions, such as smartphone or internet addiction (Choi et al., 2014; Kim et al., 2014), which were found to be related to similar levels of impulsivity as pathological gambling (Lee et al., 2012).

The objective of the present study is to test a larger sample of participants recruited in the clinical setting with the aim to compare their motor-related introspective reports (i.e. the M and W reports) to a sample of healthy control participants. The goals are to replicate results of some of the previous studies on this topic, to explore whether the effect of delayed W reports can be found in diagnoses other than those already described in the literature and to assess reliability of the Libet's clock task with the prospect to offer it as basis for a psychodiagnostical method. Besides general exploration of the tested variables, I will specifically investigate following hypotheses:

H₁: Individuals with addiction provide later W reports than controls.

> This hypothesis is based on the reported connection between impulsivity and the Libet's task (see Caspar and Cleeremans, 2015) combined with the assumption of higher impulsivity in individuals with addiction when compared to healthy individuals (see Coskunpinar et al., 2013, and other sources above).

---

[13] It should be noted that these authors showed that different facets of impulsivity trait influence alcohol addiction differently.

H₂ₐ: Individuals with neurosis provide later M reports than controls.

H₂ᵦ: Individuals with neurosis provide later W reports than controls.

The inspiration for these hypotheses originated in Edwards et al.'s (2011) findings that clinical patients with psychogenic tremor report both later M and later W judgements than healthy controls. Admittedly, the authors interpret the results in terms of decreased volitionality of the tremor and not in terms of a neurotic process. However, given the seemingly broad applicability of the later W reports also to other psychopathological categories, I attempt to generalize the findings and hypothesize that later M and W reports might be related to neurotic disorders in general.

H₃: Individuals with psychosis provide later W reports than controls.

Pirio Richardson et al. (2006) reported suggestively—although not significantly—later W reports provided by participants with schizophrenia compared to healthy controls, hence inspiring this hypothesis.

6.2 Materials and methods

6.2.1 Participants

The study was conducted on 108 participants (67 males) ranging from 17 to 74 years of age ($M = 35.19$, $SD = 12.41$), 11 left-handed. The sample consisted of a clinical and control group.

The clinical sample originally consisted of 58 individuals recruited in the Psychiatric Hospital Kroměříž (30 participants) and the Psychiatric Hospital "Marianny Oranžské" in Bílá voda (28 participants). The medical staff was informed about the research and asked to pass the information to relevant patients. With respect to the study's goals, I preferred patients with localized brain lesions (especially in the parietal area), schizophrenia, addiction or somatoform neurotic disorders. A priori contraindications were epilepsy, uncorrected sight impairment, acute psychosis and low intellect or organic disorder. All participants were currently hospitalized and were briefly informed about the research from the medical staff. Their participation was voluntary—in the case of their interest they were instructed to approach a staff member, who then mediated the contact between me and the patient. Eleven participants were excluded for the following reasons: unclear diagnoses (2 participants), dual diagnoses categorized in different diagnostical groups (5 participants), organic disorder potentially preventing correct task comprehension and execution (3 participants) and refusal

to provide informed consent (1 participant). Final number of 47 participants remained in the clinical group (30 males, in range from 17 to 61 years of age, M = 36.47, SD = 11.32). Combining information from the patients' self-reports, medical history and officially assigned diagnoses, we identified four general diagnostical groups—neurotic disorders, depression, addiction and psychotic disorders (see Table 10 for details). For a summary of participants' education background and comparison with the control group, see Table 11.

Category	Issued diagnosis		Frequency
addiction	dependence syndrome: alcohol (F10.2)		5
	dependence syndrome: stimulants (F15.2)		4
	dependence syndrome: multiple drugs (F19.2)		11
	pathological gambling (F63.0)		1
		Total:	21
depression	bipolar affective disorder, current episode mild or moderate depression (F31.3)		1
	depressive episode (F32)		2
	recurrent depressive disorder (F33)		2
		Total:	5
neurotic disorders	phobic anxiety disorders (F40)		1
	other anxiety disorders (F41)		9
	obsessive-compulsive disorder (F42)		1
	reaction to severe stress and adjustment disorders (F43)		4
	dissociative disorders (F45)		2
		Total:	17
psychotic disorders	paranoid schizophrenia (F20.0)		3
	acute polymorphic psychotic disorder without symptoms of schizophrenia (F23.1)		1
		Total:	4

Table 10: Overview of the clinical sample in Study 3 with respect to diagnostical categories. Column "category" signifies general diagnostical cluster analysed in this study. Codes in the "issued diagnosis" column designate diagnoses according to ICD-10 (World Health Organization, 1992); some categories' labels are shortened.

The control group (n = 51) was recruited from psychology students at Palacký University Olomouc (15 participants), from examinees in the Military University Hospital Prague (army employees and applicants, 26 participants) and by intentional sampling of adult participants with vocational or high education (10 participants). Eleven control participants were excluded from the main analyses for misunderstood instruction, significant fatigue during the session or self-reported psychopathology; however, some of the participants were included in the test-retest correlation analysis if found appropriate. The remaining 40 participants (28 males, in range from 18 to 56 years of age, M = 32.25, SD = 10.36)

constituted a control sample of diverse educational background and age, which at least approximately reflected the characteristics of the clinical sample (see Table 11).

Education	Clinical group 30 males, 17 females age M = 36.47, SD = 11.32	Control group 28 males, 13 females age M = 32.41, SD = 10.29
High school students	0	2
University students	3	8
Grammar school graduates	3	0
Vocational school graduates	26	8
High school graduates	12	12
University graduates	3	10

Table 11: Overview of the sample in Study 3 with respect to educational background and gender.

6.2.2 Technical equipment

Because this study pertains introspective reports only, all equipment needed was a laptop computer with a mouse, which are both easily portable. This allowed me to conveniently collect the data from psychiatric patients directly in the hospital. The equipment in absolute majority of cases[14] consisted of a laptop computer ASUS ZenBook UX305CA with cable-connected mouse Dell MS111-L.

The utilized clock software was almost identical to the one used in Studies 1 and 2 (see chapters 4.2.2 and 5.2.2 and Figure 2b in chapter 2.2.3) but was modified in three aspects. First, we converted it to an offline webpage running on the laptop locally—this was necessary to make the experiment independent on internet access. Second, we made some adjustments to the interface. Namely, we removed unnecessary options in the clock setup and left only participant's ID, gender, handedness, length of examination (10, 20 and 30 trials) and notes. Because we originally planned to provide the software to several examiners with presumably differently sized displays, we also included an option to set the clock diameter by adjusting the width of a rectangle on the screen to the width of A4 paper sheet by pressing left or right arrow keys. Even though this was not ultimately necessary—since all examinations were conducted by me personally—the idea of adjusting display setting by standardized paper sheet might be potentially useful in various future experiments. Third, compared to Studies 1 and 2 we significantly simplified the instruction to the form of several simple sentences and an animation of the running clock (see Appendix 7).

---

[14] Ten participants were tested on regular office computers, see below.

Collected reports were either M or W (see chapter 2.3.1), all made in the absolute mode of recall (see the end of chapter 2.3). The mouse could be held in either left or right hand, but almost all participants preferred the right hand (including all but one of the left-handed individuals).

6.2.3 Experimental procedure and design

To maximize comparability of the results, standardized data collection procedure was followed. All participants were interviewed in personal contact, provided informed consent and were tested in face-to-face situation (the only exception from this rule were some of the psychology students, see below). I will first introduce the general procedure and subsequently explain specifics of certain participant groups.

General procedure was as follows. After an individual decided to participate in the research—based on the information provided by the medical staff or other mediators—we arranged a term of meeting. Upon arrival, I greeted the participant and introduced myself. Subsequently, I presented the informed consent (see Appendix 8) and explained its content thoroughly. This was not only intended to treat potential ethical problems, but also to explain the research procedure and rationale in a comprehensive way. The participants confirmed consent by signing the document. The patients in the clinical group were subsequently interviewed about their education, work, current mental health issues (including their history), current or past somatic disorders, medication (if known to the participant) and length of current hospitalization. In the control group, the participants were only asked about their education, work history, health issues (including mental health) and medication. After the interview, the task was explained and ten training trials—five in the M and five in the W conditions—were conducted. If the participant clearly misunderstood the instruction, he or she was corrected during the training trials. When it was clear that the participant understands the task, 20 M and 20 W regular trials were presented. After the task ended, all data were saved, but only those from the regular trials were used in the analyses. During the whole procedure—since the beginning of the meeting until the end of the regular trials—the participant was being observed and present mental status was assessed. After the task was finished, the participant was asked what he or she thinks about the experiment and whether he or she was able to distinguish between the M and W reports. If there was enough time left and the participant was willing, 10 more trials in each condition were executed to increase

the amount of data for analysis. After the session was over, the participant was informed that he or she can contact me and ask for results, which eight participants did.

The participants in the clinical group also provided consent with my consultation of their medical records archived in the institution they were tested in. In the medical records, I verified each patient's medical history, length and circumstances of the current hospitalization and the officially issued diagnoses. I also copied the details of current medication (active ingredients, dosage and current effect, if applicable). If it was possible, I also briefly interviewed individual patients' psychologists or nurses to further clarify the information. For more information on medical records processing, see chapter 6.2.4.

An exception from the general data collection procedure explained above were participants recruited from the full-time psychology students within their cognitive psychology class. The recruitment was realized by me personally and additionally by e-mail. By this method, I recruited 10 volunteers. These participants were tested primarily for the purposes of test-retest analysis (i.e. twice), but their results were found suitable to be used in the control group as well. They were tested all at once in a computer lab at the Department of Psychology at Palacký University Olomouc. Displays of the used computers were set to correspond to the clock size presented to all other participants; computer mice were also comparable to the one used in other examinations. Because face-to-face interview was in this case impracticable, the questions were presented in the form of an open-ended questionnaire. Because I did not have any practical reason for collecting the student's names, these participants passed the experiment under anonymous codes randomly assigned at the beginning of the first session.

Fifteen participants in the control group were tested twice to assess test-retest reliability of the task. One of the participants was excluded due to reported organic disorder. Five of the remaining 14 participants reported other forms of psychopathology (mostly neurotic) but were not excluded from the test-retest analysis. The repeated session was identical to the first one with the exception that education, work, medical history and long-term medication were no longer inquired about. The time interval between the first and the second session was 8.4 days on average (SD = 3.7, median = 7, min = 6, max = 11); typical interval was 7 days (applicable for 10 participants). All other participants were tested in the standard way already described.

In some cases, extraordinary events occurred during the data collection. For instance, while testing some participants in the control group, a loud music played in neighbouring part of the building could be heard. These extraordinary circumstances were noted, but not considered within the analyses.

6.2.4 Ethics of the clinical study

Because the experiment required testing clinical patients, potential ethical issues were considered, and steps were taken to minimize the risks. The study was approved by the ethical committee at the Department of Psychology, Faculty of Arts, Palacký University Olomouc (see Appendix 9). Each participant provided an informed consent by signing a document (see Appendix 8), which contained following general points:

1. **Information about the experiment**: title, author, supervisor, grant support and general objective (i.e. development of a computerized method for mental health diagnostics)

2. Summary of **participant's task**, including the general procedure and its length

3. Information about **voluntary participation** (specifically, the participant was informed that he or she can decline participation at any time without any adverse consequences)

4. Information about the option of **retroactive consent withdrawal** by contacting me before the date of the final anonymization of all data

5. **Potential health risks** of the study, especially risk for individuals with epilepsy (given that the participants watched animated clock on a computer display)

6. Consent with **acquisition and retainment of medical and personal data**, specifically name, gender, birth date, education, work, medication and—for clinical group only—issued diagnoses and length of current disorder; this includes consent with disclosure of medical documentation

7. Detailed **explanation of data processing** procedure

8. Information about the possibility to **request personalized results**.

All participants were eligible for legal acts in the extent needed for providing informed consent.

I should further explain several aspects of the experiment, which might potentially raise ethical issues. First, there were two **reasons for collecting participants' names**. One reason was to allow me to search patients' medical documentations. From the procedural point of

view, it would be unreasonably difficult and technically inconvenient for the medical staff to issue anonymous labels to the patients and then provide me extracts from their records. The second reason for retaining participants' names was to allow them to request their personal results without having to memorize or otherwise retain a random anonymous code. For these two reasons, the names of the participants were processed together with all other data. However, all data were anonymized on the 1st January 2019, which was stated in the written consent beforehand to encourage the participants to request their results before this date.

Second potential ethical issue is the **medical records handling**. Probably the most ethically acceptable procedure would be to ask medical staff to provide extracts from the patients' documentation, but again this would burden the staff with unnecessary additional tasks. Therefore, I agreed with the medical staff superiors on two options. One option was that I reviewed the documentation myself, but only under supervision of entrusted psychologist or nurse and only within specific offices or examination rooms. Second option was that I met with entrusted psychologist and explained what information I request—the psychologist then searched for the information and read it to me aloud. Both solutions were realized; the first option in the Psychiatric Hospital in Bílá voda, the second option in the Psychiatric Hospital Kroměříž. No information was handed over via different channels, such as e-mail or phone calls, even though the patients provided consent for that as well.

Third potential issue is the **data processing and retainment**. The processing steps were as follows:

1. Task results (i.e. the M and W reports) and participants' names were saved locally immediately after the task was finished.[15]
2. My notes from the examination were transcribed to an electronic form.
3. The data from the medical documentation were provided to me in a way described above (for the clinical group only).
4. The task results, electronical notes and information from the medical records (for the clinical group only) were merged based on participants' names.

---

[15] In fact, the informed consent stated that the results will be sent to a cloud storage via internet. However, this was ultimately impracticable because of the lack of internet access in the hospital buildings. Nevertheless, because the participants agreed with transferring the results via internet, there is no reason to assume that they would not agree with its retainment in an offline storage, which is objectively safer.

5. The data were anonymized with an exception of two backup offline media (USB drives) stored physically in a safe place. These backups were retained until the 1st January 2019 for the eventuality that some participants request their personalized results or withdraw their consent.

6. All remaining data were processed and analysed anonymously.

For the anonymously tested participants recruited from the psychology students, all the described steps applied, with the difference that their data were processed under their anonymous codes instead of names.

All participants were granted the opportunity to retroactively withdraw their informed consent, which would mean deletion of their data from all storages. However, this eventuality never occurred.

6.2.5 Variables

The dependent variable in all variants of the analysis is the reported timing of the introspective events relative to the mouse click in each individual trial.[16] Within the main analysis, the timing was related to one principle independent variable based on the study's design and several covariates:

(1) **Diagnostical group.** Principle factor included to analyse the difference between the control group and four large diagnostical categories (neurotic disorders, depression, addiction and psychotic disorders); polytomous fixed factor referenced to the control group.

(2) **Trial.** Included in the analysis to account for the succession of the individual trials in the respective series; continuous fixed covariate.

(3) **Subject.** Included to consider that every 40 trials (if no trials added or excluded) were provided by the same participant; polytomous random covariate.

(4) **Gender.** Included to test whether there is a difference in response tendencies between males and females; no effect is expected. Dichotomous fixed covariate referenced to males.

(5) **Age.** Included to test whether there is a relationship between age and response tendencies; no effect is expected. Continuous fixed covariate.

---

[16] I.e. each row in our dataset represents one trial.

(6) **Education.** Included to the analysis *ad hoc* after significant effect of age on the M reports was found (see chapter 6.3); the purpose is to test whether the effect of age is mediated by education level. Polytomous fixed covariate.

(7) **Medication.** Included to account for possible confounding effects of current medication. The medication included both long-term and recently administered short-term drugs. Because the patients were administered wide range of medicaments and because the sample is still relatively small, it would be impossible to account for every active ingredient and all possible interactions. Therefore, we reduced the medication factor to nine alternative sub-covariates. One of these was a general alternative factor stating whether any of the administered drug is considered to have **sedative effect** (according to Kopeček, Brunovský, & Páleníček, 2017). The remaining eight alternative sub-covariates reflected whether any of these drug types were administered to each respective patient:

    a. **antidepressants** (e.g. clomipramine, venlafaxine, sertraline)

    b. **antiepileptics** (e.g. gabapentin, lamotrigine, pregabalin)

    c. **antihistamines** (e.g. hydroxyzine, promethazine, bisulepin)

    d. **antipsychotics, regular** (e.g. aripiprazole, quetiapine, haloperidol)

    e. **antipsychotics, atypical** (e.g. amisulpride, sulpiride, tiapride)

    f. **anxiolytics** (e.g. diazepam, alprazolam, clonazepam)

    g. **cognition-enhancing drugs** (donepezil; factor was not used, because this medication was administered to only one participant, who was additionally excluded from the study)

    h. **mood stabilizers** (sodium-valproate, lamotrigine[17])

Besides the comparison between the clinical groups and the control group, a test-retest correlation was also assessed. For this purpose, relevant trials were labelled as being conducted in a repeated session (as opposed to the trials conducted in the first or the only session). However, for the main comparative analysis, only trials from the first/only sessions were used.

---

[17] Lamotrigine can be classified as both antiepileptics and mood stabilizer. Therefore, if any participant was medicated with lamotrigine, both "antiepileptics" and "mood stabilizers" factors were scored as true.

6.2.6 Data analysis

Basic pre-processing procedures were conducted in MS Excel 2016 and R software. First, from the original number of 4980 trials, data from excluded participants and data from invalid trials were deleted from the dataset (1164 trials excluded). Second, reporting bias potentially caused by extremely late and extremely early reports was corrected, leading to correction of 24 trials (the procedure was the same as corresponding procedure described in chapter 4.2.5). Third, outliers were removed from each series based on the Tukey's 1.5×IQR rule (142 trials excluded). The remaining dataset consisted of 3674 trials.

The hypothesis that individuals in the clinical groups differ from the control group in the M and W reports was tested using two mixed-effect linear models—one for the M and one for the W reports—implemented using the R software[18]. Variables included in the models are explained in chapter 6.2.5. Unlike in Study 1, no hierarchical structure of factors was constructed, because the only absolute linear dependency is expected between the factor Subject and all other independent variables (except for Trial), which is accounted for by considering Subject a random factor.

The reliability of the M and W reports was assessed using two procedures. One procedure was based on comparing the within-subject and between-subject variability, the other procedure was test-retest correlation. The variability analysis was performed on data subsampled from the data used for the model above, excluding all clinical groups. The idea was that a method can be considered reliable if the within-subject variance is lower than between-subject variance. The within-subject variance was calculated by extracting residual variance from the M and W reports using a mixed-effect linear model including only the factor Subject (as fixed factor in this case). The between-subject variance was calculated as total variance minus the within-subject variance. The reliability of a single trial was calculated as the ratio of the between-subject variance and total variance[19], and reliability of the whole series was obtained by using the Spearman-Brown prediction formula extending the single trial reliability to a set of 20 trials.

The test-retest analysis was conducted on trials obtained from the participants who were willing and found suitable to be tested also in a repeated session (n = 14). The original sample

---

[18] Packages lme4 and lmerTest were employed. Shatterthwaite approximations to degrees of freedom were employed in significance tests of dependent variables.

[19] This procedure is based on the same principle as the analysis of variance (ANOVA).

of these paired data consisted of 1120 trials (28 M series and 28 W series from 14 participants), 47 of which were excluded due to their identification as outliers. The remaining 1073 trials were used to calculate means for individual series and these means were correlated between the sessions.

6.3 Results

The main analysis of clinical groups was performed for the M and W reports separately. In this chapter, I report results of mixed-effect linear models primarily investigating whether clinical groups differ from the control group. At the end of the chapter, reliability analyses are presented.

Factor	Estimate	t value (df)	p value (*two-tailed*)	p value (*one-tailed*)
Intercept	90.27			
Addiction	82.05	2.205 (86.40)	0.030*	
Depression	-16.91	-0.199 (85.33)	0.843	
Neurosis	-78.53	-1.202 (86.09)	0.233	0.884 (H_{2a})
Psychosis	-59.97	-0.647 (86.02)	0.519	
Gender (female)	-20.86	-0.649 (86.22)	0.518	
Age	-3.99	-2.887 (86.52)	0.005**	
Antidepressants	89.49	1.380 (86.37)	0.171	
Antiepileptics	8.02	0.126 (85.85)	0.900	
Antihistamines	97.56	1.829 (85.81)	0.071	
Antipsychotics, regular	-25.96	-0.392 (85.88)	0.696	
Antipsychotics, atypical	59.62	0.652 (85.90)	0.516	
Anxiolytics	-37.07	-0.593 (85.61)	0.555	
Mood stabilizers	-32.42	-0.276 (85.70)	0.783	
Sedative effect	8.50	0.108 (86.80)	0.914	
Trial	0.42	0.959 (1597.74)	0.338	

Table 12: Summary of the factors' effects on the M reports within the comparison of the clinical groups and the control group, while accounting for covariates of gender, age, medication, trial order and participants (participants' effects are not displayed, because the factor was considered random and hence its estimate cannot be calculated). Estimate value of the intercept designates the average M report of medication-free males of "0 age" in the control group. Single asterisk indicate significance at $\alpha = 0.05$, double asterisk indicates significance at $\alpha = 0.01$. The column "p value (one-tailed)" contains only p value related to one-tailed hypothesis H_{2a} presented in the introduction.

The analysis of the M reports showed significant main effect of addiction, $t(86.40) = 2.205$, $p = 0.030$, and highly significant main effect of age, $t(86.52) = -2.887$, $p = 0.005$. All remaining effects were not significant (see Table 12). H_{2a} about expected later M reports provided by individuals with neurosis was not supported, $t(86.09) = -1.202$, $p = 0.884$, with the M reports earlier than those provided by control group, suggesting opposite effect than expected. It should be noted that the residual distribution exhibited high level of kurtosis, possibly notably weakening power of the statistical test.

Because the age effect was unexpected, we decided to investigate it further. It can be argued that the effect of age can be mediated by education, because older participants in the sample generally tended to have vocational education, while the younger participants tended to have higher-level education achieved at high schools or universities. However, adding the factor Education did not significantly increase the amount of variance explained by the model, $F(5, 87.31) = 1.053$, $p = 0.392$, and the effect of age remained highly significant, $t(86.66) = -3.290$, $p = 0.001$.

Similar analysis was also performed for the W reports, yielding several significant effects (see Table 13). The results showed highly significant effects of addiction, $t(85.25) = 2.683$, $p = 0.009$, and neurosis, $t(84.44) = 2.784$, $p = 0.007$, generally shifting the W reports to earlier times. Therefore, the hypotheses H_1 (expecting later W reports in the addiction group) and H_{2b} (expecting later W reports in the neurosis group) were not supported—$p = 0.996$ and $p = 0.997$, respectively—because the found effects were in the opposite direction than expected. The same applies to H_3 (expecting later W reports in the psychosis group), in which the found effect was not significant and in the opposite direction as well, $p = 0.536$. Nevertheless, significant and highly significant effects were additionally found for majority of the medication (antidepressants, antiepileptics, antihistamines, anxiolytics and sedative drugs in general) and for the trial order effect. It should be noted that the atypical antipsychotics also exhibited strong practical effect, although it was not identified as significant due to limited sample of individuals medicated with atypical antipsychotics. Again, the residual distribution was found to exhibit excessive kurtosis, possibly limiting the test power.

Factor	Estimate	t value (df)	p value (*two-tailed*)	p value (*one-tailed*)
Intercept	-279.42			
Addiction	-148.85	-2.683 (85.25)	0.009**	0.996 (H₁)
Depression	-216.97	-1.722 (83.30)	0.089	
Neurosis	-269.20	-2.784 (84.44)	0.007**	0.997 (H₂ᵦ)
Psychosis	-12.60	-0.092 (85.14)	0.927	0.536 (H₃)
Gender (female)	-11.61	-0.237 (84.95)	0.813	
Age	1.88	0.882 (85.19)	0.380	
Antidepressants	318.74	3.318 (85.35)	0.001**	
Antiepileptics	222.51	2.360 (83.97)	0.021*	
Antihistamines	205.27	2.598 (84.03)	0.011*	
Antipsychotics, regular	-56.13	-0.573 (84.07)	0.568	
Antipsychotics, atypical	260.83	1.932 (83.86)	0.057	
Anxiolytics	281.15	3.041 (84.06)	0.003**	
Mood stabilizers	-4.78	-0.028 (83.25)	0.978	
Sedative effect	-387.60	-3.317 (85.93)	0.001**	
Trial	2.00	2.115 (1556.48)	0.035*	

Table 13: Summary of the factors' effects on the W reports within the comparison of the clinical groups and the control group, while accounting for covariates of gender, age, medication, trial order and participants (participants' effects are not displayed, because the factor was considered random and hence its estimate cannot be calculated). Estimate value of the intercept designates the average W report of medication-free males of "0 age" in the control group. Single asterisk indicate significance at $\alpha = 0.05$, double asterisk indicates significance at $\alpha = 0.01$. The column "p value (one-tailed)" contains only p values related to one-tailed hypotheses H_1, H_{2b} and H_3 presented in the introduction.

Because the medication effects on the W reports are in orders of hundreds of milliseconds, we performed an analysis in which we excluded all medication from the model. Interestingly, the results no longer showed significant effects of the clinical groups (see Table 14), strongly suggesting that to find a significant effect of any of the four diagnostical categories, medication needs to be considered. A χ^2-test showed significant difference of the amount of variance explained by the two models, $\chi^2(8) = 20.46$, $p > 0.001$.

Factor	Estimate	t value (df)	p value (*two-tailed*)	p value (*one-tailed*)
Intercept	-291.23			
Addiction	-107.58	-1.862 (85.63)	0.066	0.967 (H₁)
Depression	-32.68	-0.304 (84.30)	0.762	
Neurosis	-40.09	-0.649 (84.51)	0.518	0.741 (H₂ᵦ)
Psychosis	-158.84	-1.446 (85.92)	0.152	0.924 (H₃)
Gender (female)	-23.65	-0.464 (85.06)	0.644	
Age	2.35	1.068 (85.18)	0.289	
Trial	2.03	2.154 (1555.74)	0.031*	

Table 14: Summary of the factors' effects on the W reports within the comparison of the clinical groups and the control group, while accounting for covariates when the medication effect is not considered. Estimate value of the intercept designates the average W report of males of "0 age" in the control group. Single asterisk indicates significance at $\alpha = 0.05$. The column "p value (one-tailed)" contains only p values related to one-tailed hypotheses H_1, H_{2b} and H_3 presented in the introduction.

Reliability analysis was performed in two ways: by a variance analysis and by a test-retest correlation (see chapter 6.2.6). The variance analysis showed the M reports reliability of 0.898 and the W reports reliability of 0.943. Test-retest correlation for the M reports was not significant, $r(12) = 0.516$, $p = 0.059$, but highly significant for the W reports, $r(12) = 0.849$, $p < 0.001$. The underestimated reliability coefficient in the test-retest analysis might be caused by a limited sample compared to the sample used in the variance-analysis method. Generally lower reliability for the M reports might be caused by their much lower variability (SD = 158.72) compared to variability of the W reports (SD = 329.60), $F(559, 559) = 4.312$, $p < 0.001$[20].

I will now summarize the findings with regards to proposed hypotheses:

H₁: Individuals with addiction provide later W reports than controls.

Hypothesis was not accepted, $p = 0.996$. Significant effect in the opposite direction was found.

---

[20] This F-test must be considered orientational only; it neglects the assumption of independent samples and independent observations.

H$_{2a}$: Individuals with neurosis provide later M reports than controls.

Hypothesis was not accepted, p = 0.884. No significant effect was found.

H$_{2b}$: Individuals with neurosis provide later W reports than controls.

Hypothesis was not accepted, p = 0.997. Significant effect in the opposite direction was found.

H$_3$: Individuals with psychosis provide later W reports than controls.

Hypothesis was not accepted, p = 0.536. No significant effect was found.

6.4 Discussion

The aim of this study was to investigate whether it would be meaningful to use Libet's clock task as a psychodiagnostical method to distinguish between healthy individuals and individuals with psychopathology. The task has two forms, in which the participants report the time of the movement (M) or the time of wanting to move (W).

The analysis of the M reports was expected to show significant effect of the neurosis group (hypothesis H$_{2a}$). However, the effect was in fact not significant, and a trend was shown in the opposite direction than expected. The most plausible explanation seems to be that my assumption that there will be later M reports in the neurosis group was simply incorrect because of overgeneralization of the findings of Edwards et al. (2011), who only showed this effect for individuals with psychogenic tremor, i.e. very specific symptom of neurotic disorders.

Overall, the M trials were found to be rather uninformative for the clinical diagnostics; the only relevant significant effect found for the individuals with addiction was only closely significant, hence both its usability as a diagnostical measure and reliability of such result— in the context of the problem of multiple testing—can be questioned. For these reasons, I remain sceptical that the M reports can be clinically interesting.

Nevertheless, it is interesting from the exploratory point of view that an unexpected highly significant effect of age on the M reports was found, given that this was not—to my knowledge—explored in the literature. The effect shows that with every year of age the mean M report shifts to earlier time by approximately 4 ms. At the first glance, this might not seem like much, but because the estimate applies to every year of age, the implication is that 20-year olds and 50-year olds might differ in their M reports by about 120 ms on average. It is important to suggest an explanation for this. One explanation might be that this relation is

in fact mediated by the effect of education, which was unevenly distributed throughout the age groups in my sample. The education might play a role since higher education might be connected to better cognitive abilities, including introspection, and to higher willingness to introspect. However, the further analysis showed that inclusion of the education into the model did not lead to increase in explained variance, hence not clarifying the effect of age, which remained highly significant. Another explanation might be that this is somehow connected to perceptual accuracy when interoceptively sensing one's own movement. This suggestion can be supported by the general notion that the brain development reaches its peak at early 20s (see e.g. Dosenbach et al., 2010), which coincides with the age at which—according to our model—the participants' M reports are on average equal to 0, thus being absolutely precise. The suggested relation between the M reports and age should be further examined in future studies, given that in traditional Libet-style experiments (including our Studies 1 and 2) it is not usually taken into account.

Based on findings from previous research (Sirigu et al., 2004; Pirio Richardson et al., 2006; Edwards et al., 2011; Moretto et al., 2011), it can be expected that the W reports might be potentially used to distinguish between various forms of psychopathology and mental health. The analysis, results of which are summarized in Table 13, implies highly significant effects of addiction and neurosis. However, relevant hypotheses H_1, H_{2b} and H_3—predicting significantly later W reports for addiction, neurosis and psychosis groups, respectively— were not supported by the results. In fact, the effects found were all in the opposite direction than expected, generally suggesting earlier, not later, W reports for the clinical groups. In the case of neurosis, the same explanation as in the discussion of the M reports applies: it is possible that I have overgeneralized the findings from the study of psychogenic tremor (Edwards et al., 2011). The same might apply for the addiction group, because the hypothesis relevant to addiction was based on a mediation of the relationship by impulsivity; this mediation was not guaranteed to be in effect, given that many other variables might have mediated or moderated the connection. The lack of significant effect of psychosis is not surprising, as it was not found to be statistically significant in the primary source either (Pirio Richardson et al., 2006). However, Pirio Richardson and his colleagues reported suggestive trend for later W reports in the psychosis group, which is not present in the data presented here—estimated difference in the W reports between the psychosis group and the control group is mere 13 ms. This discrepancy can be explained by the fact the both Pirio Richardson

et al.'s (2006) study and the present Study 3 only obtained data from very limited sample of patients with psychosis—6 in the original study, 4 in the present study. Further research with larger sample size would be welcome.

Depression group was not expected to yield significant results and, indeed, none were found. However, a notable trend was found in the case of the W reports, with depressive individuals providing W reports earlier than controls by 217 ms. This effect was not identified as significant because of small sample size in the depression group (n = 5). Although there seems to be no theoretical reason to expect abnormal W reports provided by depressive individuals, it might be worth testing in future studies.

As shown in Table 13, addiction and neurosis groups exhibited highly significant shifts of the W reports to earlier times, contrary to my expectations. These striking findings require an explanation, which might be found when interpreting the results in the context of Study 1 and Dominik et al. (2017). To remind the reader, the main outcome of Study 1 was that it seems that the W reports are retrospectively inferred based on the perceived M times, because the participant believes the experimenter expects him or her to have an impression of wanting to move somewhen before the M time. Therefore, this phenomenon can be framed as an attempt to comply with experimenter's assumed expectations. If this interpretation is true, then it makes sense that individuals diagnosed with neurosis and individuals in addiction treatment specifically will tend to comply more than other groups. It can be speculated that neurotic patients employ the inference strategy described above but tend to exaggerate the interval by which the W judgements should precede the M reports because of the patients' effort to comply with instruction accented by their tendency towards anxiety. In the case of individuals in the addiction group, higher level of compliance is expected because patients' compliance often determines their stay in a rehabilitation programme. On the other hand, there seems to be no theoretical reason to expect such increase in compliance in the depression and psychosis groups, therefore explaining the lack of said effect in these diagnostical categories.

The highly significant result in the neurosis group can be potentially attributed to another factor, which is the increased attentiveness of neurotic patients to their somatic processes. The anxious episodes of many neurotic disorders are typically accompanied by somatic symptoms of anxiety (see e.g. American Psychiatric Association, 2013, pp. 189–233). Therefore, it could be speculated that neurotic patients report earlier times of the first

sensation of preparing to move because they tend to observe their somatic processes with greater attention than healthy individuals. This line of research might be interesting to continue.

Another noticeable effect on the W reports was found for various types of medication. These effects were not expected; I originally assumed that the M and W reports mental processing is independent of medication effects. Nevertheless, the results suggest that antidepressants, antiepileptics, antihistamines and anxiolytics tend to shift the W reports to significantly later times. This finding is immensely important, because it might explain outcomes of the previous studies (Sirigu et al., 2004; Pirio Richardson et al., 2006; Edwards et al., 2011; Moretto et al., 2011) by suggesting that the extraordinarily late W reports in the clinical groups—universally found throughout the studies—could be attributed to the effect of medication rather than of the disorder itself[21]. Table 13 also shows highly significant effect of sedative medication, which seems to make the W reports earlier. Again, this might be explained by higher compliance of patients medicated with sedatives. Another explanation might be that the sedated patients tend to report earlier W reports because of their decreased self-awareness caused by the drug, but this would presumably predict the same effect in the case of the M reports, in which no such effect was found.

The trial order effect showed increase in the W reports by 2 ms for every trial. Similar effect was found in Study 2 (see chapter 5.3.2). The effect can be potentially attributed to general mental fatigue caused by the introspective task, but further investigation of the factor would be welcome.

It is important to note that if the task is intended to be potentially used by clinical psychologists as a diagnostical tool, its contamination by the medication effect poses a serious threat for its usability. That is because psychologists using the method would need to have complete information about each client's medication, which they often do not. Moreover, if the method is used without considering the effect of medication, the diagnostical groups cease to significantly differ from each other, as shown by further analysis, results of which are presented in Table 14.

---

[21] With one exception of Sirigu et al. (2004), who also studied patients with cerebral lesion, which did not seem to cause extraordinary W reports.

Reliability of the method was assessed using two independent procedures: by a test-retest correlation and by a within- and between-subject variance analysis. The test-retest correlation seemed to lead to underestimated value in the case of the M reports (r = 0.516), which might be due to combination of small test-retest sample size and low overall variability of the M reports. On the other hand, the variance analysis—with much higher sample size—showed satisfying level of the M reports variability (r = 0.898). Reliability of the W reports was satisfyingly supported by both the variance analysis and the test-retest correlation, r = 0.943 and r = 0.849, respectively. It should be noted that the reliability calculation based on the variance analysis—although suggestively convincing—should be accepted with caution, because the procedure assumes independence of trials within individual participants, which is violated in our data. Therefore, this reliability measure should be understood as an upper estimate.

Overall, it seems that the W reports might prove reliable and useful for a clinical diagnostic of addiction and neurotic disorders, but it has several significant downsides, such as possible mediation of the effect by participant's compliance, contamination by the medication and insufficient specificity while distinguishing between different diagnostical categories (e.g. between neurotic disorders and tendency to addiction). It might be worth suggesting an alternative, which might be similar but more appropriate in these regards. The alternative I would like to propose is the intentional binding, which is a method used for assessing implicit sense of agency (see chapter 3.5). The intentional binding effect was found to be connected to mental health by multiple studies. Haggard, Martin, Taylor-Clarke, Jeannerod, & Franck (2003) showed that patients suffering from schizophrenia tend to exhibit unusually strong intentional binding effect. In their experiment, participant's action was followed by a stimulus in a fixed interval of 250 ms. By a group of patients with schizophrenia, it was perceived as approximately 51 ms long on average, while healthy controls judged the same interval to be 229 ms long. The authors speculated that this might be caused by generally incorrect way of attributing outcomes to own actions characteristic for some psychotic disorders, such as schizophrenia. This finding was further developed by Voss et al. (2010), who showed that in healthy individuals the intentional binding effect occurs to some extent even when the outcome (a tone in this case) is expected but does not occur; nevertheless, patients suffering from schizophrenia did not exhibit this phenomenon at all, suggesting that the schizophrenia makes sense of agency dependent on the

retrospective evidence only. Using the intentional binding procedure rather than traditional Libet's clock task might bring promising results and eventually lead to devising a functional tool for diagnostic of psychotic disorders. That is because intentional binding is inherently implicit and thus it is less likely that the results would be mediated by the amount of willingness of the tested individuals to report the expected results, which seems to be the case with the W reports.

6.5 Limitations and recommendations

The present study contains several limitations, many of which were already discussed (e.g. contamination of the W reports by administered medication or small sample size in the psychosis and depression clinical groups). However, additional limitations should be noted.

Given that medication turned out to be a crucial factor influencing the W reports, it needs to be better accounted for in the future studies. In the present study, I have treated various forms of medication as covariates, attempting to eliminate the W reports variance explained by them. This procedure usually works well to isolate effect of other factors; however, in this case, there is an expectable co-occurrence between medication and psychiatric diagnosis, because no control participants took any psychoactive medication. Therefore, the effect of medication and the effect of clinical diagnosis might not be appropriately separated.

Because the specialists potentially using the method would often need to diagnose patients before their admission to clinical care, it might be advisable to standardize and norm the diagnostical method on patients before they are medicated. This might pose an extremely difficult task, given that some patients (such as those in acute psychotic episode) might be unable to be tested using the proposed method. It might be interesting, however, to administer the method in a form of a large-scale population screening while prospectively following the participants and subsequently identifying those who develop a mental health issue.

It might be argued that a difference in the task's results between the clinical groups and the control group is expected, because the two groups were tested in vastly different contexts and environments. Given that I have suggested that the W reports might be affected by participant's compliance with the examiner, this is certainly possible and might be interesting to study in future research.

There are several variables, which I have collected from the participants, but have not included in the analysis (such as length of hospitalization, medication dosage, medication

active ingredients, detailed diagnoses, somatic comorbidities etc.). This is especially important for the detailed diagnoses, because some diagnostical categories—such as the neurotic disorders—consist of various diagnoses of different etiology and symptoms. These additional factors were not included in the analyses because of our relatively limited sample. To study all said variables, a sample of several hundreds and potentially thousands of participants would be needed, because the diversity of clinical population would require overwhelming number of degrees of freedom.

One could also ask why only the absolute mode of recall was employed, given that the order mode of recall was identified as more "participant-friendly" in Study 2. The reason is that the reports based on the order mode of recall do not allow us to calculate any measure of variability, therefore making it impossible to perform any statistical analysis, regressive or otherwise.

6.6 Conclusions

The W task might become a promising foundation for a psychodiagnostical tool for certain mental health problems. However, its theoretical and methodological limitations are substantial. The method is biased by medication administered to the participants and its results do not correspond with theoretical expectations. Future research is suggested in the direction of implicit sense of agency measures, such as intentional binding procedure.

7 Summary of key conclusions of Studies 1, 2 and 3

The present thesis consisted of three studies of several aspects of Libet's experiment. **Study 1** investigated whether the introspective reports of wanting to move (W task) used in Libet-style experiments might be inferred based on the introspective impression of the actual movement (M task), rather than directly introspectively perceived. The results suggest so. If the participants are not familiar with the M task, they tend to provide W reports which seem to be identical to the M reports. In contrast, if the W reports are required after the M task is finished, participants tend to provide W reports preceding their M reports, ultimately replicating results of traditional Libet-style experiments.

Study 2 was a complex replication of the original Libet's experiment. The objectives of the study were to verify that the results remain identical in principle if the original methodological directions are followed closely, to identify potential methodological problems, to suggest solutions for said problems and to provide authentic data sample to the

scientific public. The objectives were fulfilled. In general, the results replicate Libet's key findings, but substantial differences were found in some of the results' categories, such as the introspective reports timing or the number of found event-related potentials.

Study 3 investigated whether the introspective tasks present in Libet's experiment, and especially the W task, may be utilized as a basis for a psychodiagnostical tool. Several findings in the literature suggest so. The results indeed suggest that the W task bears some potential to identify individuals with certain diagnoses, namely in the domains of neurotic disorders and addiction, but the results are inconclusive for several reasons. First, the effects found were in the opposite direction than what the literature suggested, and second, the effect was notably contaminated by medication administered to the participants in the clinical groups. Several improvements to the design were suggested and caution is advised when attempting to generalize the findings of Study 3.

Summary

Libet's experiment (Libet et al., 1982; Libet, Gleason, et al., 1983; Libet, Wright, et al., 1983; Libet, 1985) constituted a seminal research in the scientific studies of conscious control over motor activity. Benjamin Libet started his career under supervision of Dr. Bertram Feinstein, who inspired Libet to study temporal factors of human conscious experiences. Ultimately, Libet started studying human volition using a combination of cognitive methods and electrophysiological recordings, especially recordings of so-called readiness potential (RP), which is an event-related potential shown to precede a voluntary movement by a substantial amount of time (Kornhuber & Deecke, 1965).

Libet states that the main objective for his experiment was to *"compare the time of onset of the conscious intention to act and the time of onset of associated cerebral processes"* (Libet, 1985, p. 530). He obtained temporal onsets of conscious mental events by asking his participants (6 university students) to use rapidly revolving clock (called Libet's clock in this thesis) to report subjective timing of specific introspective experiences. The experiences were (1) the judgements of when the participant made a specific movement (M reports), (2) the judgements of when the participant registered the first intention ("wanting") to make the movement (W reports) and (3) the judgements of when a stimulus was delivered in a special type of tasks (S reports). The "associated cerebral processes" were obtained by Libet and his team using an electroencephalograph (EEG); Libet was particularly interested in the RPs, i.e. slow rises of negativity preceding the onset of each movement (which was measured objectively using an electromyograph). Besides the tasks already described, Libet also studied situations in which participants were asked to perform a movement when the Libet's clock reached certain position (so called P tasks), as well as situations in which the participants were supposed to prepare to make the movement just as in the P tasks, but then "veto" the movement at the last instance.

The main result of Libet's experiment showed that the RP onset generally precedes the W reports, suggesting that neural motor-related activity is present before the individual becomes conscious of a motor preparation. Nevertheless, the "veto" aspect of the experiment also suggests that the individual might still have some control over the course of the action by stopping the motor preparation process during its last 200 ms.

Libet's experiment was heavily discussed in the literature soon after its publication and remains so to this day. The most important points of the discussions are as follows:

- The W reports might be artificially induced and can be retrospectively biased (Breitmeyer, 1985; Gomes, 1998; Pockett & Purdy, 2011; Lau et al., 2007; Banks & Isham, 2009).

- The M reports seem to be notably robust (Pockett & Miller, 2007).

- Neural basis for the attention towards the W reports was found in the pre-SMA and for the intention to move itself in the posterior parietal cortex (Lau et al., 2004, 2006; Sirigu et al., 2004, Desmurget et al., 2009).

- Neural basis for the M reports was found in the cingulate motor area (Lau et al., 2004).

- It is still discussed whether Libet's clock method might be subjected to biases due to the flash-lag effect and prior entry effect (Cairney, 1975; Breitmeyer, 2002; Klein, 2002; Kawohl & Habermeyer, 2007; Pockett & Miller, 2007; Matsuhashi & Hallett, 2008; Papanicolaou, 2017).

- An alternative to the Libet's clock method based on random probing for participants' intention to act was proposed (Matsuhashi & Hallett, 2008; Verbaarschot et al., 2019).

- Lateralized readiness potentials (LRPs) instead of regular RPs were proposed for interpretation of Libet's findings (Haggard & Eimer, 1999; Trevena & Miller, 2002).

- Patterns of motor preparation occurring even several seconds before the movement were found using an fMRI (Soon et al., 2008).

- Results compatible with Libet's outcome were found on a single-neuron level (Fried et al., 2011).

- An RP-like wave can be found even when no overt movement is present (Miller et al., 2011; Alexander et al., 2016).

- An RP re-interpretation based on stochastic neural fluctuations was proposed (Schurger et al., 2012) and is currently being investigated (Jo et al., 2013; Alexander et al., 2016; Schmidt et al., 2016; Schultze-Kraft et al., 2016; Khalighinejad et al., 2018).

The empirical part of the present thesis consists of three studies. **Study 1** is based on the suggestion that the W reports might be induced artificially by the experimental instruction. If that is so, then we hypothesized that the participants infer the times reported as the W judgements based on their impression of the M time. Therefore, we investigated the W reports provided before and after the M task with no preceding training. We analysed introspective data of 20 participants, who performed the M and W tasks. The participants were divided into two groups based on the order of the tasks (group A performed the tasks in order W-M, group B in order M-W). The results showed significant interaction effect supporting our hypothesis, $t(20.02) = -3.488$, $p = 0.002$, showing that the W reports made after the M task were earlier by 275 ms than the W reports made before the M task. Interestingly, when investigating the M and W reports provided in the first session, no significant difference was found, suggesting that the difference observed in previous studies is indeed at least partially due to participants' training combined with their effort to comply with the instruction. The conclusion is that to obtain different M and W reports, a training is necessary, but also that the W reports might not be valid in the sense of Libet's interpretations.

Study 2 was a complex replication study investigating closely the methodology proposed by Libet. We conducted the original experiment with only a few necessary and extensively discussed modifications. We tested 8 participants using the Libet's clock to collect introspective reports, EEG to collect data on participants' readiness potentials and P300 waveforms, EMG to objectively measure muscle activity and a tactile stimulator to perform the S tasks. The results differed slightly in some categories from the original experiment, but the main points of Libet's interpretations were found to remain valid, especially that the W reports are preceded by the RP onsets by a notable amount of time.

Study 3 explored the possibility of using the introspective reports obtained using Libet's methodology as a psychodiagnostical tool. The study included 108 participants from both clinical and non-clinical setting, who were tested in the M and W tasks. The results showed no clinically interesting conclusions for the M task, but interesting implications for the W task in the case of the addiction and neurosis group. However, the results contradicted previous findings and seem significantly contaminated by the effect of psychoactive medication.

References

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A. L., Wheatley, T., & Tse, P. U. (2016). Readiness potentials driven by non-motoric processes. *Consciousness and Cognition, 39*, 38–47. doi:10.1016/j.concog.2015.11.011

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5.* Arlington, VA, USA: New School Library.

Andreassi, J. L. (2007). *Psychophysiology: Human Behavior and Physiological Response (5th ed.).* New York, NY, USA: Psychology Press.

Antonietti, A. (2011). Introspecting a conscious decision or the consciousness of a decision? *Consciousness and Cognition, 20*, 1916–1917. doi:10.1016/j.concog.2011.04.011

Association for the Scientific Study of Consciousness. (2019). *ASSC 22.* Retrieved February 27, 2019, from Association for the Scientific Study of Consciousness: https://theassc.org/assc-22/

Baldo, M. V., Kihara, A. H., & Namba, J. (2002). Evidence for an attentional component of the perceptual misalignment between moving and flashing stimuli. *Perception, 31*, 17–30.

Banks, W. P., & Isham, E. A. (2009). We infer rather than perceive the moment we decided to act. *Psychological Science, 4*, 387–391. doi:10.1111/j.1467-9280.2008.02254.x

Bareš, M. (2011). Kognitivní evokované potenciály. *Česká a slovenská neurologie a neurochirurgie, 74/107*(5), 508–517.

Barratt, E. S. (1983). The biological basis of impulsiveness: the significance of timing and rhythm disorders. *Personality and Individual Differences, 4*(4), 387–391. doi:10.1016/0191-8869(83)90004-1

Batthyany, A. (2009). Mental Causation and Free Will after Libet and Soon: Reclaiming Conscious Agency. In A. Batthyany, & A. Elitzur, *Irreducibly Conscious. Selected Papers on Consciousness.* Heidelberg, Germany: Universitätsverlag Winter GmbH. Retrieved March 20, 2019, from https://philarchive.org/archive/BATMCA

BBC Radio 4. (2014, November 7). *The Libet Experiment: Is Free Will Just an Illusion?* Retrieved November 1, 2015, from YouTube: https://youtu.be/OjCt-L0Ph5o

BioSemi. (n.d.). *BioSemi Products*. Retrieved May 16, 2019, from BioSemi: https://www.biosemi.com/products.htm

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience, 11*(5), 551–559. doi:10.1162/089892999563607

Brass, M., & Haggard, P. (2007). To Do or Not to Do: The Neural Signature of Self-Control. *The Journal of Neuroscience, 27*(34), 9141–9145. doi:10.1523/jneurosci.0924-07.2007

Breitmeyer, B. G. (1985). Problems with the psychophysics of intention. *The Behavioral and Brain Sciences, 8*(4), 539–540.

Breitmeyer, B. G. (2002). In Support of Pockett's Critique of Libet's Studies of the Time Course of Consciousness. *Consciousness and Cognition, 11*, 280–283. doi:10.1006/ccog.2002.0555

Cairney, P. T. (1975). The complication experiment uncomplicated. *Perception, 4*, 255–265.

Caspar, E. A., & Cleeremans, A. (2015). "Free will": are we all equal? A dynamical perspective of the conscious intention to move. *Neuroscience of Consciousness*, 1–10. doi:10.1093/nc/niv009

Coskunpinar, A., Dir, A. L., & Cyders, M. A. (2013). Multidimensionality in Impulsivity and Alcohol Use: A Meta-Analysis Using the UPPS Model of Impulsivity. *Alcoholism: Clinical and Experiemntal Research, 37*(9), 1441–1450. doi:10.1111/acer.12131

Danquah, A. N., Farrell, M. J., & O'Boyle, D. J. (2008). Biases in the subjective timing of perceptual events: Libet et al. (1983) revisited. *Consciousness and Cognition, 17*, 616–627.

Danto, A. C. (1985). Consciousness and motor control. *The Behavioral and Brain Sciences, 8*(4), 540–541.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement Intention After Parietal Cortex Stimulation in Humans. *Science, 324*, 811–813. doi:10.1126/science.1169896

Dewey, J. A., & Knoblich, G. (2014). Do Implicit and Explicit Measures of the Sense of Agency Measure the Same Thing? *PLOS ONE, 9*(10), e110118. doi:10.1371/journal.pone.0110118

Dominik, T. (2016). The experimental verification of validity of the introspective element in Benjamin Libet's research (Unpublished diploma thesis). Olomouc, Czech Republic: Palacký University Olomouc.

Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2017). Libet's Experiment: Questioning the Validity of Measuring the Urge to Move. *Consciousness and Cognition, 49*, 255–263. doi:10.1016/j.concog.2017.01.017

Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2018a). Libet's Experiment: A Complex Replication. *Consciousness and Cognition, 65*, 1–26. doi:10.1016/j.concog.2018.07.004

Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2018b). EEG Data and Introspective Reports from the Libet's Experiment Replication. *Data in Brief, 20*, 2040–2044. doi:10.1016/j.concog.2018.07.004

Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., . . . Coal, R. S. (2010). Prediction of Individual Brain Maturity Using fMRI. *Science, 329*, 1358–1361. doi:10.1126/science.1194144

Doty, R. W. (1985). The time course of conscious processing: Vetoes by the uninformed? *The Behavioral and Brain Sciences, 8*(4), 541–542.

Edwards, M. J., Moretto, G., Schwingenschuh, P., Katschnig, P., Bhatiaa, K. P., & Haggard, P. (2011). Abnormal sense of intention preceding voluntary movement in patients with psychogenic tremor. *Neuropsychologia, 49*, 2971–2973. doi:10.1016/j.neuropsychologia.2011.05.021

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. doi:10.3758/BF03209424

Ersche, K. D., Turton, A. J., Pradhan, S., Bullmore, E. T., & Robbins, T. W. (2010). Drug Addiction Endophenotypes: Impulsive Versus Sensation-Seeking Personality Traits. *Biological Psychiatry, 68*(8), 770–773. doi:10.1016/j.biopsych.2010.06.015

Forster, B. B., MacKay, A. L., Whittall, K. P., Kiehl, K. A., Smith, A. M., Hare, R. D., & Liddle, P. F. (1998). Functional magnetic resonance imaging: the basics of blood-oxygen-level dependent (BOLD) imaging. *Canadian Association of Radiologists Journal, 49*(5), 320–329.

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition. *Neuron, 69*, 548–562. doi:10.1016/j.neuron.2010.11.045

Gomes, G. (1998). The Timing of Conscious Experience: A Critical Review and Reinterpretation of Libet's Research. *Consciousness and Cognition, 7*, 559–595.

Gomes, G. (2002). Problems in the Timing of Conscious Experience. *Consciousness and Cognition, 7*, 191–197.

Guggisberg, A. G., Dalal, S. S., Schnider, A., & Nagarajan, S. S. (2011). The neural basis of event-time introspection. *Consciousness and Cognition, 20*, 1899–1915. doi:10.1016/j.concog.2011.03.008

Haas, L. (2003). Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery, and Psychiatry, 74*(1), 9. doi:10.1136/jnnp.74.1.9

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research, 126*, 128–133.

Haggard, P., & Libet, B. (2001). Conscious Intention and Brain Activity. *Journal of Consciousness Studies, 8*(11), 47–63.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382–385. doi:10.1038/nn827

Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. (2003). Awareness of action in schizophrenia. *Cognitive Neuroscience and Neuropsychology, 14*(7), 1081–1085. doi:10.1097/01.wnr.0000073684.00308.c0

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An Online Neural Substrate for Sense of Agency. *Cerebral Cortex, 23*(5), 1031–1037. doi:10.1093/cercor/bhs059

Choi, J.-S., Park, S. M., Roh, M.-S., Lee, J.-Y., Park, C.-B., Hwang, Y., . . . Jung, H. Y. (2014). Dysfunctional inhibitory control and impulsivity in Internet addiction. *Psychiatry Research, 215*(2), 424–428. doi:10.1016/j.psychres.2013.12.001

Jo, H.-G., Hinterberger, T., Wittmann, M., Borghardt, T. L., & Schmidt, S. (2013). Spontaneous EEG fluctuations determine the readiness potential: is preconscious brain activation a preparation process to move? *Experimental Brain Research, 231*, 495–500. doi:10.1007/s00221-013-3713-z

Kawohl, W., & Habermeyer, E. (2007). Free will: Reconciling German Civil Law with Libet studies on the Readiness Potential. *Behavioral Sciences and the Law, 25*, 309–320.

Keller, I., & Heckhausen, H. (1990). Readiness Potentials Preceding Spontaneous Motor Acts - Voluntary vs Involuntary Control. *Electroencephalography and Clinical Neurophysiology, 76*, 351–361.

Khalighinejad, N., Schurger, A., Desantis, A., Zmigrod, L., & Haggard, P. (2018). Precursor processes of human self-initiated action. *NeuroImage, 165*, 35–47. doi:10.1016/j.neuroimage.2017.09.057

Kim, S. M., Huh, H. J., Cho, H., Kwon, M., Choi, J. H., Ahn, H. J., . . . Kim, D. J. (2014). The Effect of Depression, Impulsivity, and Resilience on Smartphone Addiction in University Students. *Journal of Korean Neuropsychiatric Association, 53*(4), 214–220. doi:10.4306/jknpa.2014.53.4.214

Klein, S. (2002). Libet's Research on the Timing of Conscious Intention to Act: A Commentary. *Consciousness and Cognition, 11*, 273–279. doi:10.1006/ccog.2002.0557

Kopeček, M., Brunovský, M., & Páleníček, T. (2017). *Klinická psychofarmakologie.* (P. Mohr, Ed.) Prague, Czech Republic: Maxdorf.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv, 284*, 1–17.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On Measuring the Perceived Onsets of Spontaneous Actions. *The Journal of Neuroscience, 26*(27), 7265–7271. doi:10.1523/jneurosci.1138-06.2006

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2007). Manipulating the Experienced Onset of Intention after Action Execution. *Journal of Cognitive Neuroscience, 19*(1), 81–90. doi:10.1162/jocn.2007.19.1.81

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to Intention. *Science, 303*(5661), 1208–1210. doi:10.1126/science.1090973

Lee, H. W., Choi, J.-S., Shin, Y.-C., Lee, J.-Y., Jung, H. Y., & Kwon, J. S. (2012). Impulsivity in Internet Addiction: A Comparison with Pathological Gambling. *Cyberpsychology, Behavior, and Social Networking, 15*(7), 373–377. doi:10.1089/cyber.2012.0063

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences, 8*, 529–539.

Libet, B. (2000). Time Factors in Conscious Processes: Reply to Gilberto Gomes. *Consciousness and Cognition, 9*, 1–12. doi:10.1006/ccog.1999.0408

Libet, B. (2004). *Mind Time.* Cambridge, MA, USA: Harvard University Press.

Libet, B., Alberts, W. W., Wright, E. W., & Feinstein, B. (1967). Responses of Human Somatosensory Cortex to Stimuli below Threshold for Conscious Sensation. *Science, 158*, 1597–1600.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential); the unconscious initiation of a freely voluntary act. *Brain, 106*, 623–642.

Libet, B., Wright, E. W., & Gleason, C. A. (1982). Readiness-potentials preceding unrestricted 'spontaneous' vs. pre-planned voluntary acts. *Electroencephalography and clinical Neurophysiology, 54*, 322–335.

Libet, B., Wright, E. W., & Gleason, C. A. (1983). Preparation- gold intent-to-act, in relation to pre-event potentials Recorded at the vertex. *Electroencephalography and Clinical Neurophysiology, 56*(4), 367–372.

Lush, P., & Dienes, Z. (2019). Time perception and the experience of agency in meditation and hypnosis. *PsyCh Journal, 8*, 36–50. doi:10.1002/pchj.276

Lush, P., Naish, P., & Dienes, Z. (2019). Metacognition of intentions in mindfulness and hypnosis. *Neuroscience ofConsciousness, 1*, 1–10. doi:10.1093/nc/niw007

Ly, A., Verhagen, J., & Wagenmakers, E. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology, 72*, 19–32. doi:10.1016/j.jmp.2015.06.004

MacKay, D. M. (1958). Perceptual Stability of a Stroboscopically Lit Visual Field containing Self-Luminous Objects. *Nature, 181*, 507–508.

Matsuhashi, M., & Hallett, M. (2008). The timing of the conscious intention to move. *European Journal of Neuroscience, 28*, 2344–2351. doi:10.1111/j.1460-9568.2008.06525.x

Mill, J. S. (1882). *A System of Logic, Ratiocinative and Inductive.* New York, NY, USA: Harper & Brothers. Retrieved March 17, 2019, from https://www.gutenberg.org/files/27942/27942-h/27942-h.html#toc57

Miller, J., Shepherdson, P., & Trevena, J. (2011). Effects of Clock Monitoring on Electroencephalographic Activity: Is Unconscious Movement Initiation an Artifact of the Clock? *Psychological Science, 22*(1), 103–109. doi:10.1177/0956797610391100

Miller, J., Vieweg, P., Kruize, N., & McLea, B. (2010). Subjective reports of stimulus, response, and decision times. *Consciousness and Cognition, 19*, 1013–1036.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition, 21*(1), 546–561. doi:10.1016/j.concog.2011.12.002

Moore, J. W., Middleton, D., Haggard, P., & Fletcher, P. C. (2012). Exploring implicit and explicit aspects of sense of agency. *Consciousness and Cognition, 21*, 1748–1753. doi:10.1016/j.concog.2012.10.005

Moretto, G., Schwingenschuh, P., Katschnig, P., Bhatia, K. P., & Haggard, P. (2011). Delayed experience of volition in Gilles de la Tourette syndrome. *Journal of Neurology, Neurosurgery & Psychiatry, 82*, 1324–1327. doi:10.1136/jnnp.2010.221143

Nielsen, D. A., Ho, A., Bahl, A., Varma, P., Kellogg, S., Borg, L., & Kreek, M. J. (2012). Former heroin addicts with or without a history of cocaine dependence are more impulsive than controls. *Drug and Alcohol Dependence, 124*(1–2), 113–120. doi:10.1016/j.drugalcdep.2011.12.022

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256–257.

Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review, 84*(3), 231–259.

Obhi, S. S., Planetta, P. J., & Scantlebury, J. (2009). On the signals underlying conscious awareness of action. *Cognition, 110*, 65–73. doi:10.1016/j.cognition.2008.10.009

Papanicolaou, A. C. (2017). The Myth of the Neuroscience of Will. *Psychology of Consciousness: Theory, Research, and Practice, 4*(3), 310–320. doi:10.1037/cns0000116

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*(4), 456–479.

Pirio Richardson, S., Matsuhashi, M., Voon, V., Peckham, E., Nahab, F., Mari, Z., . . . Hallett, M. (2006). Timing of the sense of volition in patients with schizophrenia. *Clinical Neurophysiology, 117*, S97. doi:10.1016/j.clinph.2006.06.120

Pockett, S., & Miller, A. (2007). The rotating spot method of timing subjective events. *Consciousness and Cognition, 16*(2), 241–254. doi:10.1016/j.concog.2006.09.002

Pockett, S., & Purdy, S. (2011). Are voluntary movements initiated preconsciously? The relationships between readiness potentials, urges and decisions. In W. Sinnott-Armstrong, & L. Nadel, *Conscious Will and Responsibility: A Tribute to Benjamin Libet* (pp. 34–46). New York, NY, USA: Oxford University Press. doi:10.1093/acprof:oso/9780195381641.001.0001

Robert, C. P., Chopin, N., & Rousseau, J. (2009). Harold Jeffreys's Theory of Probability Revisited. *Statistical Science, 24*(2), 141–172. doi:10.1214/09-STS284

Sanquist, T. F., Beatty, J. T., & Lindsley, D. B. (1981). Slow potential shifts of human brain during forewarned reaction. *Electroencephalography and Clinical Neurophysiology, 51*(6), 639–649. doi:10.1016/0013-4694(81)90208-X

Shrout, P. E. (1998). Measurement reliability and agreement in. *Statistical Methods in Medical Research, 7*, 301–307.

Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Tse, P. U., & Wheatley, T. (2015). Hypnotizing Libet: Readiness potentials with non-conscious volition. *Consciousness and Cognition, 33*, 196–203.

Schmidt, S., Jo, H.-G., Wittmann, M., & Hinterberger, T. (2016). 'Catching the waves' – slow cortical potentials as moderator of voluntary action. *Neuroscience and Biobehavioral Reviews, 68*, 639–650. doi:10.1016/j.neubiorev.2016.06.023

Schott, G. D. (1993). Penfield's homunculus: a note on cerebral cartography. *Jounal of Neurology, Neurosurgery, and Psychiatry, 56*, 329–333.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., . . . Haynes, J.-D. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences of the United States of America, 113*(4), 1080–1085. doi:10.1073/pnas.1513569112

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), E2904–E2913.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience, 7*(1), 80–84. doi:10.1038/nn1160

Smallwood, J., & Schooler, J. W. (2006). The Restless Mind. *Psychological Bulletin, 132*(6), 946–958. doi:10.1037/0033-2909.132.6.946

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Science, 11*(5), 543–545. doi:10.1038/nn.2112

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological Recording.* New York, NY, USA: Oxford University Press.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 629–685). New York, NY, USA: Academic Press.

Trans Cranial Technologies. (2012). *Cortical Functions.* Hong Kong, China: Trans Cranial Technologies ldt.

Trevena, J. A., & Miller, J. (2002). Cortical Movement Preparation before and after a Conscious Decision to Move. *Consciousness and Cognition, 11*, 162–190.

Trevena, J., & Miller, J. (2010). Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition, 19*, 447–456. doi:10.1016/j.concog.2009.08.006

van Boxtel, G. J., & Brunia, C. H. (1994). Motor and non-motor aspects of slow brain potentials. *Biological Psychology, 38*, 37–51. doi:10.1016/0301-0511(94)90048-5

van den Broek, M. D., Bradshaw, C. M., & Szabadi, E. (1992). Performance of impulsive and non-impulsive subjects on two temporal differentiation tasks. *Personality and Individual Differences, 13*(2), 169–174. doi:10.1016/0191-8869(92)90039-R

Vasser, M., Vuillaume, L., Cleeremans, A., & Aru, J. (2019). Waving goodbye to contrast: self-generated hand movements attenuate visual sensitivity. *Neuroscience ofConsciousness, 5*(1), 1–9. doi:10.1093/nc/niy013

Verbaarschot, C., Farquhar, J., & Haselager, P. (2015). Lost in time... The search for intentions and Readiness Potentials. *Consciousness and Cognition, 33*, 300–315. doi:10.1016/j.concog.2015.01.011

Verbaarschot, C., Haselager, P., & Farquhar, J. (2019). Probing for Intentions: Why Clocks Do Not Provide the Only Measurement of Time. *Frontiers in Human Neuroscience, 13*, 1–20. doi:10.3389/fnhum.2019.00068

Verleger, R., Haake, M., Baur, A., & Śmigasiewicz, K. (2016). Time to Move Again: Does the Bereitschaftspotential Covary with Demands on Internal Timing? *Frontiers in Human Neuroscience, 10*(642), 1–13. doi:10.3389/fnhum.2016.00642

Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain, 133*(10), 3104–3112. doi:10.1093/brain/awq152

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain. *Nature, 203*, 380–384.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*(1), 231–240.

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. *Cognition, 121*(2), 207–218. doi:10.1016/j.cognition.2011.06.011

Wittmann, M., Leland, D. S., Churan, J., & Paulus, M. P. (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug and Alcohol Dependence, 90*, 183–192. doi:10.1016/j.drugalcdep.2007.03.005

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva, Switzerland: World Health Organization.

Appendices

Appendix 1: Abstract of the thesis (EN)

Title: Libet's experiment, its replicability, validity and clinical potential

Author: Mgr. Tomáš Dominik

Supervisor: PhDr. Mgr. Roman Procházka, Ph.D.

Number of pages and characters: 151 pages, 351 542 characters

Number of appendices: 9

Number of references: 111

Abstract:

The present thesis can be divided into four parts, which are theoretical introduction and three empirical studies. In the theoretical part, the reader is introduced to the principle of Libet's experiment, which studied neural and introspective component of human volition, and to many of the discussions pertaining it. In Study 1, the validity of Libet's method of measuring the intention to move (W) is studied by asking the participants to provide timing of the W and of the M (time of the actual movement) in two opposite orders (W-M and M-W). The results showed that the W reports might be artificially induced by the experimental instruction suggesting that intention should precede the movement. Study 2 was a replication study, in which the original experiment was reproduced with only a few modifications. The results confirmed the original conclusions, although outcomes in some categories differed from the original experiment. In Study 3, potential of the M and W reports for psychological diagnostics was examined by testing clinical patients and healthy controls. The results suggest some potential, but exhibited large effects of confounding variables, such as medication.

Keywords: Libet's experiment, voluntary action, subjective events timing, rotating spot method, readiness potential

Appendix 2: Abstrakt disertační práce (CZ)

Název: Libetův experiment, jeho replikabilita, validita a klinický potenciál

Autor: Mgr. Tomáš Dominik

Školitel: PhDr. Mgr. Roman Procházka, Ph.D.

Počet stran a znaků: 151 stran, 351 542 znaků

Počet příloh: 9

Počet titulů použité literatury: 111

Abstrakt:

Tuto práci lze rozdělit do čtyř částí, kterými jsou teoretický úvod a tři empirické studie. V teoretické části je čtenář seznámen s principem Libetova experimentu, který zkoumal neurální a introspektivní složku lidské vůle, a také s mnoha diskuzemi, které se experimentu týkají. Ve Studii 1 je zkoumána validita Libetovy metody měření záměru k pohybu (W). Participanti měli za úkol uvádět čas W nebo M (což je čas pohybu samotného), a to v jednom ze dvou pořadí (W-M a M-W). Výsledky ukázaly, že reporty W mohou být uměle indukovány experimentální instrukcí naznačující, že záměr by měl předcházet pohybu. Ve Studii 2 byl původní experiment reprodukován, a to jen s malým množstvím modifikací. Výsledky potvrdily původní závěry, ačkoli zjištění v některých kategoriích se od původního experimentu lišily. Ve Studii 3 byl zkoumán potenciál M a W reportů pro psychologickou diagnostiku. Byli testování kliničtí pacienti a zdraví účastníci. Výsledky naznačují určitý potenciál, avšak byl zjištěn výrazný vliv dalších proměnných, jako je například medikace.

Klíčová slova: Libetův experiment, volní akce, subjektivní časování, rotating spot method, readiness potential

Appendix 3: Co-authors' declaration

I agree with the use of the following papers as part of the doctoral thesis of Mgr. Tomáš Dominik titled *"Libet's experiment, its replicability, validity and clinical potential"*:

- Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2017). Libet's Experiment: Questioning the Validity of Measuring the Urge to Move. *Consciouness and Cognition, 49*, 255-263. doi:10.1016/j.concog.2017.01.017
- Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2018a). Libet's Experiment: A Complex Replication. *Consciousness and Cognition, 65*, 1-26. doi:10.1016/j.concog.2018.07.004
- Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2018b). EEG Data and Introspective Reports from the Libet's Experiment Replication. *Data in Brief, 20*, 2040-2044. doi:10.1016/j.concog.2018.07.004

I recognize Mgr. Tomáš Dominik as the principle author of these papers and give him full permission to reprint their text in his thesis and to change the text accordingly.

PhDr. Daniel Dostál, Ph.D.   Mgr. et Mgr. Martin Zielina, Ph.D.
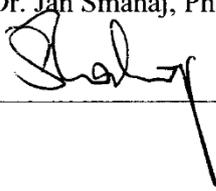
PhDr. Jan Šmahaj, Ph.D.   PhDr. Mgr. Roman Procházka, Ph.D.
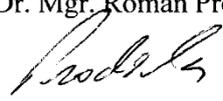
Mgr. Zuzana Vaculčíková Sedláčková, Ph.D.

Appendix 4: Experimental instruction used in Study 1

This appendix provides an overview of the electronic instructions which were presented to the participants in Study 1. The instructions were presented in Czech and slightly varied depending on experimental conditions, their order and subject's handedness (alternatives presented in italics, only one of which was displayed to the participant) and gender (adding "-a" to the end of verbs in the case that the participant was female). In this translation, I present four versions of the instruction:

1. First 40 trials in group A (W condition)
2. Second 40 trials in group A (M condition)
3. First 40 trials in group B (M condition)
4. Second 40 trials in group B (W condition)

Note that the sentences in the instructions below may feel somewhat unnatural. This might be caused by the translation procedure which attempts to remain close to the original Czech version, but also to follow the rules of English as closely as possible.

The instruction and the image of the clock face were white on black background.

Appendix 4.1, group A, first 40 trials (W condition):

"Lay your hand comfortably on the table and place your *right index finger on the right Enter key / left index finger on the left Ctrl key*. On the screen, you will see a clock face and two white spots. The first spot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second spot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second spot. For a better idea, look at the image at the bottom of this page. This clock will be the same during the whole experiment. While the clock is revolving, keep the cursor out of the clock face, so that it does not distract you.

Your task will be to let the clock complete one whole revolution and then press the key in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means, do not plan your movement. Make the movement instantaneously, just as you feel the urge to do it.

After the keypress, the spot on the circumference will keep moving for a while. Then the computer will ask you which time the clock was showing in the moment you realized the first urge to press the key. Insert the answer value by left-clicking on a certain spot on the clock. To relieve your eyes, we recommend blinking a few times and focus your gaze off the screen after some answers. Then click on the OK button and continue with another trial. Keep in mind that you report the value afterwards, so that you do not have to feel distracted while pressing the key. After the series is finished, another instruction will appear. You may use the pause to briefly relax. Then continue by reading the new instruction.

In the phase of pressing the key, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before going on. Remember that while making the keypress, you have to keep your eyes focused on the white spot in the middle of the clock face."

Long press the key to continue…

Appendix 4.2, group A, second 40 trials (M condition):

"The following part of the experiment will be similar to the previous. Again, your task is to press the key at whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means, do not plan your movement. Make the movement instantaneously, just as you feel the urge to do it.

The difference is as follows: after the keypress, the spot on the circumference will keep moving for a while. Then the computer will ask you which time the clock was showing in the moment you realized that you are moving your finger. Insert the answer value by left-clicking on a certain spot on the clock. To relieve your eyes, we recommend blinking a few times and focus your gaze off the screen after some answers. Then click on the OK button and continue with another trial. After finishing the series, press F11 to leave full screen mode. You may then watch a video discussing this experiment.

In the phase of pressing the key, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before going on. Remember that while making the keypress, you have to keep your eyes focused on the white spot in the middle of the clock face."

Long press the key to continue…

Appendix 4.3, group B, first 40 trials (M condition):

"Lay your hand comfortably on the table and place your *right index finger on the right Enter key / left index finger on the left Ctrl key*. On the screen, you will see a clock face and two white spots. The first spot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second spot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second spot. For a better idea, look at the image at the bottom of this page. This clock will be the same during the whole experiment. While the clock is revolving, keep the cursor out of the clock face, so that it does not distract you.

Your task will be to let the clock complete one whole revolution and then press the key in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means, do not plan your movement. Make the movement instantaneously, just as you feel the urge to do it.

After the keypress, the spot on the circumference will keep moving for a while. Then the computer will ask you which time the clock was showing in the moment you realized that you are moving your finger. Insert the answer value by left-clicking on a certain spot on the clock. To relieve your eyes, we recommend blinking a few times and focus your gaze off the screen after some answers. Then click on the OK button and continue with another trial. Keep in mind that you report the value afterwards, so that you do not have to feel distracted while pressing the key. After the series is finished, another instruction will appear. You may use the pause to briefly relax. Then continue by reading the new instruction.

In the phase of pressing the key, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before going on. Remember that while making the keypress, you have to keep your eyes focused on the white spot in the middle of the clock face."

Long press the key to continue…

Appendix 4.4, group B, second 40 trials (W condition):

"The following part of the experiment will be similar to the previous. Again, your task is to press the key at whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means, do not plan your movement. Make the movement instantaneously, just as you feel the urge to do it.

The difference is as follows: after the keypress, the spot on the circumference will keep moving for a while. Then the computer will ask you which time the clock was showing in the moment you realized the first urge to press the key. Insert the answer value by left-clicking on a certain spot on the clock. To relieve your eyes, we recommend blinking a few times and focus your gaze off the screen after some answers. Then click on the OK button and continue with another trial. After finishing the series, press F11 to leave full screen mode. You may then watch a video discussing this experiment.

In the phase of pressing the key, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before going on. Remember that while making the keypress, you have to keep your eyes focused on the white spot in the middle of the clock face."

Long press the key to continue…

Appendix 5: Experimental instruction used in Study 2

This appendix provides an overview of the electronic instructions which were presented to the participants prior to each series in Study 2. The instructions were presented in Czech and were adjusted to the participant's gender (adding "-a" to the end of verbs in the case that the participant was female). Otherwise, the participants were presented the same instruction in the corresponding series.

The instructions were series-specific; i.e. the instruction differed for the M(A), M(O), W(A), W(O), S(A), S(O), P and Pv series. The same instruction was displayed twice—once before the 10 training trials and once before the 40 regular trials. The instructions were rather long, but participants were required to read and interpret them in their own words during the first few sessions. In later sessions, the participants were still encouraged to read the instructions to re-familiarize themselves with the procedure but were no longer required to interpret them.

Note that the sentences in the instructions below may feel somewhat unnatural. This might be caused by the translation procedure which attempts to remain close to the original Czech version, but also to follow the rules of English as closely as possible.

The instruction and the image of the clock face were white on black background.

Appendix 5.1, M(A) series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to let the clock complete one whole revolution and then press the mouse button in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means that you should not plan your movement, and rather move instantaneously, just as you feel the urge to do it.

After the mouse click, the dot on the circumference will keep moving for a while and then the computer will ask you what the clock was showing when you realized that you are **moving** your finger. Insert the answer value by left-clicking on the corresponding spot on the clock. To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen after some answers. Then click on the OK button and continue with another trial.

Keep in mind that you report the value afterwards, hence you do not have to feel distracted while pressing the key.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before continuing. Remember that while making the click, you must keep your eyes focused on the white dot in the middle of the clock face.

Press and hold the mouse button to continue…

Appendix 5.2, M(O) series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to let the clock complete one whole revolution and then press the mouse button in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means that you should not plan your movement, and rather move instantaneously, just as you feel the urge to do it.

After the mouse click, the dot on the circumference will keep moving for a while and then jump to a certain position. The computer will ask you whether, when you realized that you are **moving** your finger, the dot on the circumference showed earlier, the same or later time in comparison with what time the clock will be currently showing. Insert the answer value by left-clicking on the corresponding answer button. You do not have to remember this, as the buttons will always contain labels as well. After the answer, the computer will continue with the next trial.

To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen while the clock is running. After you regain concentration, wait at least one whole revolution before continuing.

Keep in mind that you report the value afterwards, hence you do not have to feel distracted while pressing the key.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before continuing. Remember that while making the click, you must keep your eyes focused on the white dot in the middle of the clock face.

Press and hold the mouse button to continue…

Appendix 5.3, W(A) series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to let the clock complete one whole revolution and then press the mouse button in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means that you should not plan your movement, and rather move instantaneously, just as you feel the urge to do it.

After the mouse click, the dot on the circumference will keep moving for a while and then the computer will ask you what time the clock was showing when you realized the first **urge** to press the mouse button. Insert the answer value by left-clicking on the corresponding spot on the clock. To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen after some answers. Then click on the OK button and continue with another trial.

Keep in mind that you report the value afterwards, hence you do not have to feel distracted while pressing the key.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before continuing. Remember that while making the click, you must keep your eyes focused on the white dot in the middle of the clock face.

Press and hold the mouse button to continue…

Appendix 5.4, W(O) series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to let the clock complete one whole revolution and then press the mouse button in whichever moment you want. Try to make the movement as fast as possible and spontaneously. This means that you should not plan your movement, and rather move instantaneously, just as you feel the urge to do it.

After the mouse click, the dot on the circumference will keep moving for a while and then jump to a certain position. The computer will ask you whether, when you realized the first **urge** to press the mouse button, the dot on the circumference showed earlier, the same or later time in comparison with what time the clock will be currently showing. Insert the answer value by left-clicking on the corresponding answer button. You do not have to remember this, as the buttons will always contain labels as well. After the answer, the computer will continue with the next trial.

To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen while the clock is running. After you regain concentration, wait at least one whole revolution before continuing.

Keep in mind that you report the value afterwards, hence you do not have to feel distracted while pressing the key.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. If the urge to blink starts to be uncomfortable, you can blink, but then wait at least one whole revolution before continuing. Remember that while making the click, you must keep your eyes focused on the white dot in the middle of the clock face.

Press and hold the mouse button to continue…

Appendix 5.5, S(A) series

In this experiment, we will not observe your hand movement, but a perception of a touch.

Sit comfortably. You have a stimulator attached to your left wrist. The stimulator will mildly tap your skin in random times, which you have already experienced. Lay both your hand comfortably on the armrests.

On the screen, you will see a familiar clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

Your task will be just to sit calmly and watch the clock.

After the tap on your skin, the dot on the circumference will keep moving for a while and then the computer will ask you what time the clock was showing when you realized the **tap**. Insert the answer value by left-clicking on the corresponding spot on the clock. To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen after some answers. Then click on the OK button and continue with another trial.

In the phase of waiting for the tap, try not to blink. Remember that you must keep your eyes focused on the white dot in the middle of the clock face.
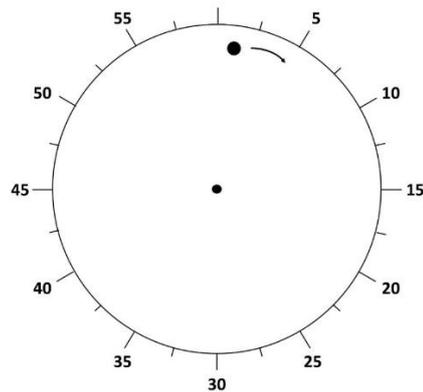
Press and hold the mouse button to continue…

Appendix 5.6, S(O) series

In this experiment, we will not observe your hand movement, but a perception of a touch.

Sit comfortably. You have a stimulator attached to your left wrist. The stimulator will mildly tap your skin in random times, which you have already experienced. Lay both your hand comfortably on the armrests.

On the screen, you will see a familiar clock face and two white dots. The first dot will be displayed in the middle of the clock face—focus your gaze on it and do not take your eyes off. The second dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. Do not focus your gaze on the second dot. Remember to keep the cursor out of the clock face, so that it does not cover it.

Your task will be just to sit calmly and watch the clock.

After the tap on your skin, the dot on the circumference will keep moving for a while and then jump to a certain position. The computer will ask you whether, when you realized the **tap**, the dot on the circumference showed earlier, the same or later time in comparison with what time the clock will be currently showing. Insert the answer value by left-clicking on the corresponding answer button. You do not have to remember this, as the buttons will always contain labels as well. After the answer, the computer will continue with the next trial.

To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen while the clock is running. After you regain concentration, wait at least one whole revolution before continuing.

In the phase of waiting for the tap, try not to blink. Remember that you must keep your eyes focused on the white dot in the middle of the clock face.
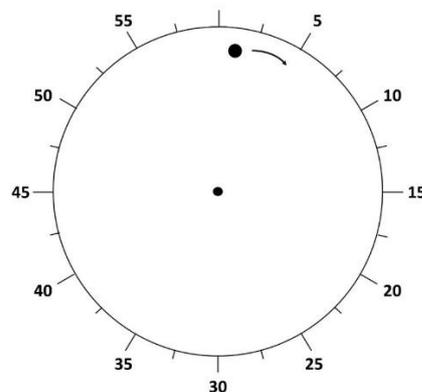
Press and hold the mouse button to continue…

Appendix 5.7, P series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots and one green dot in it. Focus your gaze on the white dot in the middle of the clock face and do not take your eyes off. The second white dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. The green dot will be placed somewhere on the circumference and will remain stationary. Do not focus your gaze on the dots on the circumference. Also remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to **<u>click</u>** the mouse button in the precise moment when the white dot is at the same position as the green dot. Try to make the movement as fast as possible. A few moments after the click, the green dot will disappear, and after a while re-appear somewhere else. Your task will then be repeated. In this experiment, you will not state any reports on the clock.

If you miss the moment marked by the green dot, wait calmly for the white dot to complete the current revolution and then try again. The computer will not manipulate the green dot unless you click the mouse button. To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen after each click. After you regain concentration, wait at least one whole revolution before continuing.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. Remember that you must keep your eyes focused on the white dot in the middle of the clock face.

Press and hold the mouse button to continue…

167
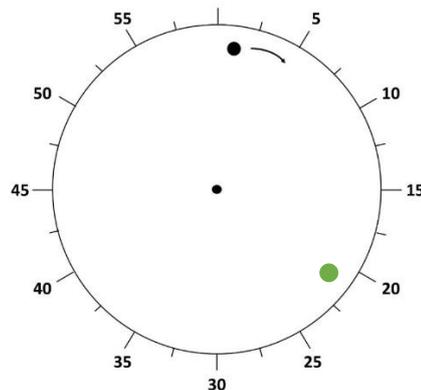
Appendix 5.8, Pv series

Lay your hand comfortably on the mouse.

On the screen, you will see a clock face and two white dots and one green dot in it. Focus your gaze on the white dot in the middle of the clock face and do not take your eyes off. The second white dot will revolve around the clock face in such a distance, that you can comfortably watch it peripherally. The green dot will be placed somewhere on the circumference and will remain stationary. Do not focus your gaze on the dots on the circumference. Also remember to keep the cursor out of the clock face, so that it does not cover it.

For a better idea what the clock will look like, look at the image at the bottom of this page.

Your task will be to **prepare** to click the mouse button in the precise moment when the white dot is at the same position as the green dot but stop the movement in the last moment. It is important that you try to prevent even a little movement of the hand muscles.

Click the mouse button whenever during the following revolution of the clock to inform the computer that you "stopped" of the movement. A few moments after the click, the green dot will disappear, and after a while re-appear somewhere else. Your task will then be repeated. In this experiment, you will not state any reports on the clock.

If you miss the moment marked by the green dot, wait calmly for the white dot to complete the current revolution and try again. The computer will not manipulate the green dot unless you click the mouse button at some time during the clock running. To relieve your eyes, we recommend blinking a few times and focusing your gaze off the screen while the clock is running (but not between stopping the movement and the click). After you regain concentration, wait at least one whole revolution before continuing.

After the series is finished, another instruction will appear.

In the phase of pressing the mouse button, try not to blink. Remember that you must keep your eyes focused on the white dot in the middle of the clock face.
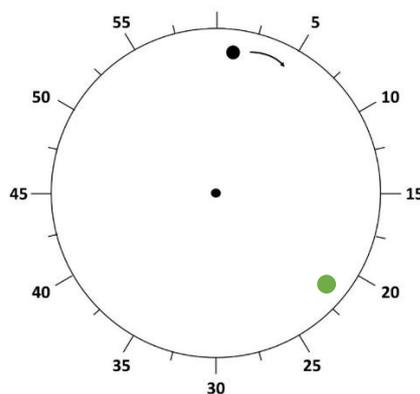
Press and hold the mouse button to continue…

# Appendix 6: Guide to the RP identification used in Study 2

This document was used as a guideline for the eye-ball RP identification and onset assessment Study 2. It is not meant to represent any official guideline; its purpose was specific to present study only. However, we encourage the scientific community to devise a similar, universally valid guideline to standardize RP identification using an eye-ball inspection.

This guideline does contain some features that were later included in the research. For example, in its Section 2 the Guide states that an RP is not supposed to be found where the RP negativity stagnates for more than 500 ms; this rule was held less strictly in our research, because it led to omission of many type I RPs. Note that Section 3 of this appendix contains technical directions which refer to other files not included in this appendix.

To summarize, this guideline was used as a tentative tool and needs improvements before being offered as a generally functional instrument.

Appendix 6, section 1: Observed waveforms characteristics

- RP characteristics:
    - negative shift preceding the time 0
    - should not start later than 50 ms before the movement
    - should have an amplitude about -0,01 mV (our Y axis is in millivolts) – if the amplitude is pronouncedly lower (e.g. 0.001 mV) or larger (e.g. 0.03 mV), we do not consider it an RP
    - typically, the RP is of one of three types (images are idealized depiction containing no noise):
        1. RP I – has a ramp-like form (long slope) and starts usually more than 700 ms before the movement

        0,01 mV

        1000 ms
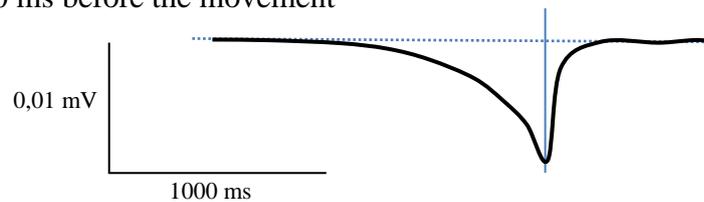
        2. RP II – has a dome-shaped form and starts between 700 and 400 ms before the movement; RP II can be occasionally preceded by a small negativity wave

        0,01 mV

        1000 ms
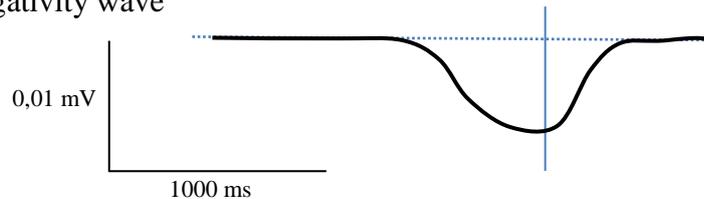
        3. RP III – starts between 200 and 250 ms before the movement; it is short and shallow


        0,01

        1000 ms
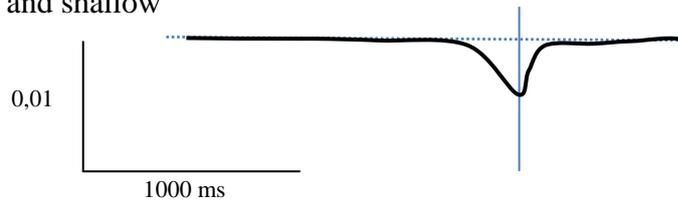
        4. pre-set RP – starts early, has a large amplitude and irregular shape; may begin returning to baseline even a few hundred ms before reaching the "0 time"


        0,01

        1000 ms
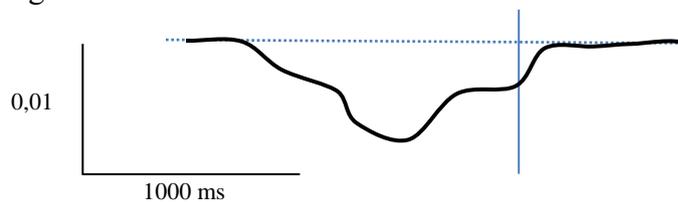
- P300 wave:
  - in some series we also expect a P300 waveform complex
  - P300 is a distinct massive "hill" of positivity with a peak at about **300 ms after** an onset of a stimulus expected by the participant
  - P300 wave is usually preceded by wave-like complex with usually lower amplitude
 - idealized depiction:
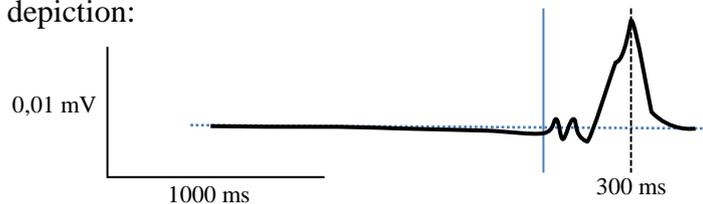
Appendix 6, section 2: Assessment criteria

- we judge the EEG plot starting at about -1800 ms (we do not expect any wave to initiate before this point)
- we assess following three characteristics:
  1. RP onset
  2. RP type
  3. P300 occurrence
- RP onset is assessed as follows:
  - the onset should be placed at a time when the wave starts to decline into negativity **beyond the noise variance** -> i.e. the RP onset is placed (with regards to the voltage on Y axis) approximately on the mean level of preceding baseline, not on the highest peak of the noise (see the following picture):
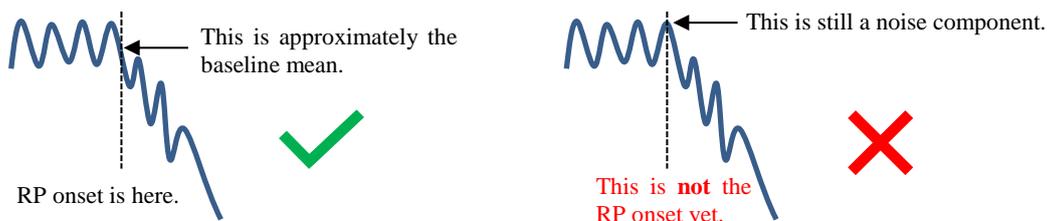

  - RP onset is not where:
    1. the decline stops and remains at the same level for more than 500 ms
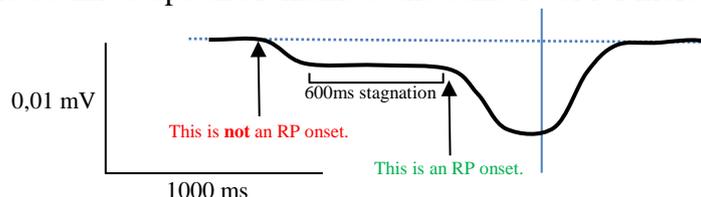
    2. the curve returns to baseline **more than 200 ms** before the movement (does not apply for the pre-set RP)


171
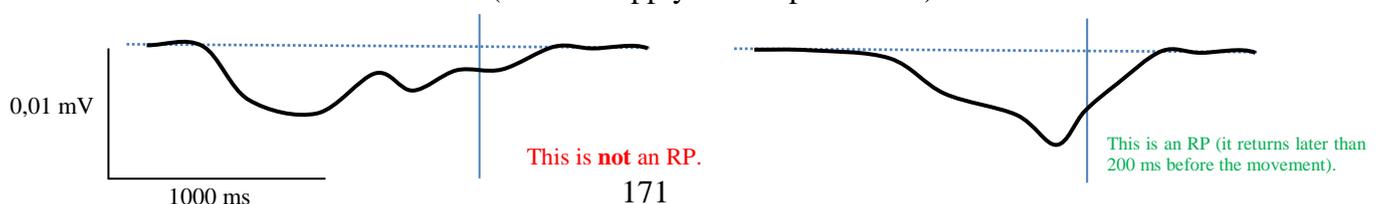
3. the curve returns to baseline after a positive shift
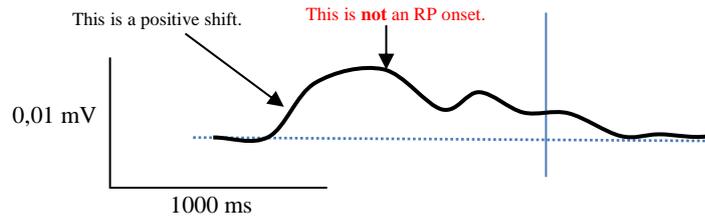
4. negativity does not return to baseline after the movement and continues to decline into negative values.
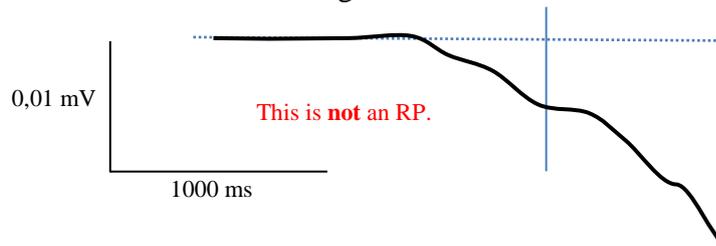
- P300 waveform is assessed as follows:
  - we judge only whether we see it or not (nothing else need to be inferred)
  - it can happen that in one recording both an RP and P300 is found—they are not mutually exclusive
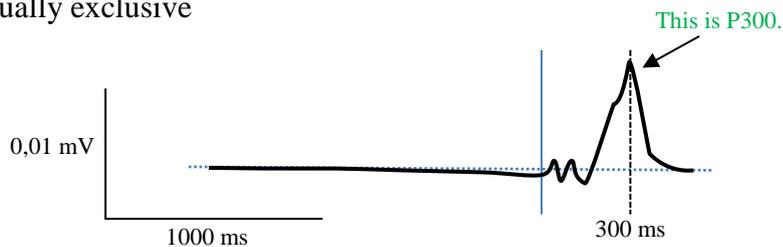
- CAUTION: It can also happen that a plot does not contain any RP or P300—in such a case, we simply do not report anything

Appendix 6, section 3: Assessment procedure

1) Open the file *RP spreadsheet.xlsx* (our working document, not present here).
2) Click on "1" in the first column under the header → plot 1 will open in a web browser.
3) **Check the scale** on the Y axis (so that you know how large the wave amplitude is).
4) Assess, whether **an RP is present** (see the rules in Section 2).
5) Assess the **RP onset** (when pointing the cursor somewhere on the plot, a timestamp and a voltage in mV will appear in the top right corner).
6) Enter the RP onset into the column *RP onset* (if an RP is not detected, leave the cell blank)
7) Assess the **RP type** (mainly according to shape, but also with respect to its size and onset; see Section 1) and enter the RP type with a number (1, 2 or 3) into the column *RP type*; if you cannot decide or an RP is missing, leave the cell blank.

8) Assess whether **a P300 is present**—if so, enter "1" into the *P300* column; if not, leave the cell blank.

9) **Enter any notes**, ideas and problems into the *note* column (problems stated in Section 2 do not have to be reported)

10) In the plot, click on the "Následující" (Czech for "Next") and repeat.

11) Take a break after every 100 plots.

TIPS:

a) If you want to return to an interrupted work, just open the RP spreadsheet and click on the number of the plot you want to work on.

b) Zooming – by holding the left mouse button and dragging over the plot, you can zoom the view (horizontally and vertically, independently on each other); you can return to the default view by double-clicking the left mouse button.

Appendix 7: Experimental instruction used in Study 3

This appendix provides an overview of the electronic instructions which were presented to the participants prior to each series in Study 3. The instructions were presented in Czech and slightly varied depending on experimental conditions, their order and subject's handedness (alternatives presented in italics, only one of which was displayed to the participant) and gender (adding "-a" to the end of verbs in case the participant was female). Otherwise, the participants were presented the same instruction in the corresponding series, as well as their oral interpretation provided by me. In this translation, I present four instructions, which were always presented in this order:

1. 5 trials in the M condition (training)
2. 5 trials in the W condition (training)
3. 40 trials in the M condition (regular)
4. 40 trials in the W condition (regular)

Note that the sentences in the instructions below may feel somewhat unnatural. This might be caused by the translation procedure which attempts to remain close to the original Czech version, but also to follow the rules of English as closely as possible. Unlike in previous studies, instruction in Study 3 is notably shorter to allow for easier explanation to all clinical patients. Also, the instruction in Study 3 presented the clock face as an animation.

The instruction and the animation of the clock face were white on black background.

Appendix 7.1, training M series:

There are two tasks ahead. You will first practice both and then perform them one more time.

Lay you *right/left* hand on the mouse.

On the screen, you will see a clock face with a cross in the middle and moving dot on the circumference:
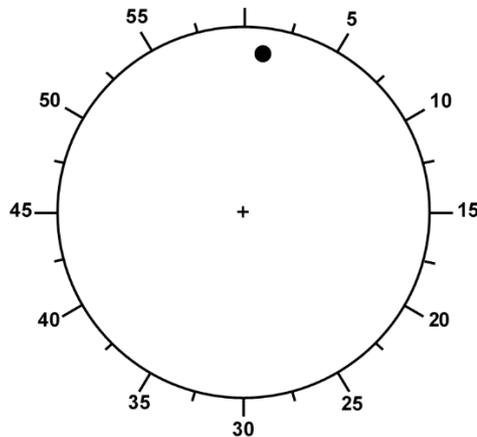
Your task will be as follows:

1. Fix your gaze on the cross in the middle of the clock face.
2. During the first revolution, do nothing.
3. Then, while still watching the cross, click the *left/right* mouse button whenever you feel like it.
4. After a while, the computer will ask **where the dot was in the moment you realized that your finger moved**. Answer by clicking on the corresponding position on the clock face.
5. Continue by clicking on the OK button.

While performing the mouse click, do not blink. **The click should be sudden, and you should not plan it.**

Please, explain now to the examiner what your task is.

Appendix 7.2, training W series:

Well done! We will now continue.

This task will be almost the same.

Once again, lay your *right/left* hand on the mouse.

On the screen, you will see the same clock face with a cross in the middle and moving dot on the circumference:
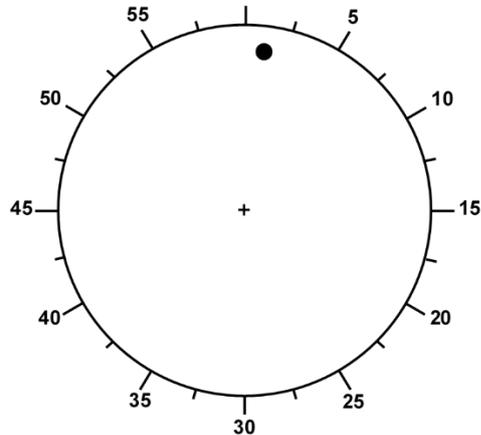
Your task will the same as in previous case, with **one difference in step 4**:

1. Fix your gaze on the cross in the middle of the clock face.
2. During the first revolution, do nothing.
3. Then, while still watching the cross, click the *left/right* mouse button whenever you feel like it.
4. After a while, the computer will ask **where the dot was in the moment you realized that you are about to move your finger**. Answer by clicking on the corresponding position on the clock face.
5. Continue by clicking on the OK button.

While performing the mouse click, do not blink. **The click should be sudden, and you should not plan it.**

Please, explain now to the examiner what your task is and how it differs from the previous task.

Appendix 7.3, regular M series:

You are now familiar with the whole task. We will now repeat the tasks and will start with the first variant (watch when you finger moved):

1. Fix your gaze on the cross in the middle of the clock face.
2. During the first revolution, do nothing.
3. Then, while still watching the cross, click the *left/right* mouse button whenever you feel like it.
4. After a while, the computer will ask **where the dot was in the moment you realized that your finger moved**. Answer by clicking on the corresponding position on the clock face.
5. Continue by clicking on the OK button.

While performing the mouse click, do not blink. **The click should be sudden, and you should not plan it.**

Appendix 7.4, regular W series:

We will now repeat the other task (watch when you are about to move your finger):

1. Fix your gaze on the cross in the middle of the clock face.
2. During the first revolution, do nothing.
3. Then, while still watching the cross, click the *left/right* mouse button whenever you feel like it.
4. After a while, the computer will ask **where the dot was in the moment you realized that you are about to move your finger**. Answer by clicking on the corresponding position on the clock face.
5. Continue by clicking on the OK button.

While performing the mouse click, do not blink. **The click should be sudden, and you should not plan it.**

Appendix 8: Informed consent in Study 3

Note: this informed consent was provided in Czech and differed slightly between the clinical and control group; specifically, the informed consent for the control group did not contain some irrelevant information, such as consent with disclosure of medical records. This Appendix presents the informed consent for the clinical group.

Informed consent with participation in a research

Dear Sir or Madame,

you were approached with an offer to participate on a scientific project called *Libet's experiment in the clinical context*. The project is realized by Mgr. Tomáš Dominik (further referred to as the "author of the study" or the "author") with support of a grant IGA 2017 and under supervision of PhDr. Mgr. Roman Procházka, Ph.D. and the Department of Psychology, Faculty of Arts, Palacký University Olomouc. The project is aimed on developing a new psychodiagnostical method based on measuring how humans experience movement of their muscles. Research so far suggests that such a measurement can give us information about some aspects of people's psychological health.

The task is to **perform several very simple operations**, which will be explained on the screen. Specifically, you will repeatedly click a mouse button while watching a special clock face and after each click, you will state when you registered certain psychological phenomena connected to the finger movement. **The task lasts 10–30 minutes**. Your **participation is voluntary**, and you have full right to state your refusal to participate by not signing this document. Your refusal with further participation will have no undesirable effects.

If you decide to participate in the study and to sign this document, you still retain **your right to contact the author of the study with request to retroactively withdraw consent and request deletion of all your data**, which you can do until when it can no longer be determined which data belong to you (i.e. until the 1st January 2019, see below). However, you express your consent that the study with your anonymous data included can be published even before this date.

It is highly unlikely that your participation in the study poses any health risk for you. However, if you have ever experienced an epileptic episode (seizure) and think that there is a risk of its another occurrence while watching computer screen, by signing this document you express your consent that neither the author of the study nor the institution within which you are being examined are responsible for this potential event.

By your signature you further express consent that the institution within which you are examined provides information from your medical records to the author, specifically: your name and surname, your gender, your birth date, your education, information about your medication, potential issued diagnoses and length of potential disorder. This information, together with the results from the testing procedure, will be processed in the following steps:

1. Automatic transfer of the task's results and your name to the author immediately after finishing the task.
2. Transfer of the stated information from your medical records from the institution within which you are examined to the author, either electronically or in person.
3. Pairing of the task results and your medical information based on your name.
4. **Anonymization** of the paired **data**, i.e. removing your name and transforming your birth date to age in years.
5. **Archiving** of the paired **non-anonymous data** on a safe electronical offline storage until the 1st January 2019. On this day, your non-anonymous data will be permanently deleted. Archiving will be done to allow the author to remove your data if you decide to withdraw your consent (see above).
6. Statistical processing of the anonymous data and their use in author's dissertation thesis and other potential scientific publications of the author.

In the case of your further interest, your participation in the study gives you full right to **contact the author of the study and request a report with your results from the task**. It is of course necessary to state your name and surname in the request, so that the author can identify your data. The author will provide the report within 30 days by e-mail.

In a case of further questions regarding the study or your rights, please contact the author on the following address: tomas.dominik01@upol.cz.

Place and date

I sincerely thank you for your participation!

Mgr. Tomáš Dominik
Ph.D. student
Palacký University Olomouc
Faculty of Arts | **Department of Psychology**
tomas.dominik01@upol.cz | www.upol.cz

Your name and surname (legibly)

Your signature

Appendix 9: Approval of Study 3 by the ethical committee at the Department of Psychology, FF UP (CZ)

KATEDRA PSΨCHOLOGIE
FILOZOFICKÁ FAKULTA
UNIVERZITA PALACKÉHO V OLOMOUCI

Korespondenční adresa: Křížkovského 10, 771 80 Olomouc
Sídlo: Vodární 6, 779 00 Olomouc
Tel.: +420 585 633 501 I Fax: +420 585 633 700
Email: psychologie@upol.cz I www.psych.upol.cz

Vyjádření etické komise Katedry psychologie FF UP v Olomouci

Etická komise Katedry psychologie FF UP v Olomouci projednala plán výzkumného projektu *Libetův experiment v klinickém kontextu*, který předložil student doktorského studijního programu klinická psychologie Mgr. Tomáš Dominik pod vedením školitele PhDr. Mgr. Romana Procházky, Ph.D. Komise prodiskutovala etické aspekty navrhovaného výzkumu a shledala, že **projekt je plně v souladu s požadavky Etického kodexu EFPA** (Evropské federace psychologických asociací).

V Olomouci dne 31. 10. 2017

Předsedkyně etické komise: PhDr. Olga Pechová, Ph.D.

Podpis:

Místopředseda etické komise: PhDr. Miroslav Charvát, Ph.D.

Podpis:

Člen etické komise: PhDr. Daniel Dostál, Ph.D.

Podpis: