VOLITION AND INHIBITION:

OBJECTIVE AND SUBJECTIVE ASPECTS OF HUMAN VOLITIONAL CONTROL

THESIS SUBMITTED IN FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

Elisa Filevich

Supervisor: Prof Patrick Haggard

Second Supervisor: Prof Paul Burgess

Institute of Cognitive Neuroscience, University College London

December 2012

Declaration

I, Elisa Filevich confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis

London, December 2012

Acknowledgements

First and foremost, I would like to thank my supervisor, Patrick Haggard. It has been an inspiring privilege to work with such an insightful, incisive and supportive mentor. Like Gandalf in The Lord of the Rings (Tolkien, 1954), Patrick was somebody I knew I could count on in difficult times. Even if, from time to time, he was away fighting bigger evils. It takes wisdom and kindness to be a Gandalf, and I would like to thank him for that. He was patient and supportive towards every aspect of my PhD, and I have greatly enjoyed working with and learning from him.

I also thank my parents and the Argentinean state for having funded twenty-one years of excellent and free education. I thank the British state for funding my overseas tuition fees, and allowing me the opportunity to be in a place as culturally and scientifically interesting as London. I thank the Wellcome Trust for being so generous and open-minded in their attitude towards science funding. I thank the technical and administrative staffs in the ICN, who have always done everything in their capacities to make our stay here enjoyable and productive. I am especially grateful to Jorge, Prad, Nahid and Eamonn for their time, effort and thoughtful comments on previous versions of this thesis.

I would also like to thank all the people I have encountered along the way, that were interested in learning about my research on “free will”. Talking about my PhD with friends was inspiring and encouraging, and fun. I thank all my family for their wonderful support, encouragement, trust, patience, love and many other things. I especially thank my brother Oski for his everlasting scientific enthusiasm, and for explaining to me what energy is.

And I thank the brain for being so interesting!

Contributions

Some of the work included in this thesis has been done in collaboration with other researchers. I am grateful to the following researchers and students.

The meta-analyses on neuroimaging data reported in chapter 1 were conducted jointly with Simone Kühn. The ideas related to intentional inhibition presented in the literature review were discussed and developed jointly with Patrick Haggard and Simone Kühn.

Jim Parkinson wrote the code for both experimental tasks and part of the analysis in the experiments reported in chapter 2. I thank James Elsey for the help in testing the participants for experiment 1.

I thank Ryusuke Kakigi for providing the prototype for the itch electrodes used for the experiment reported in chapter 3.

I am also grateful to Javier Elkin, for testing participants in the experiment reported in chapter 5, and to Alexander Browning for testing participants for the experiment reported in chapter 7.

Finally, testing for the experiment reported in chapter 8 was done jointly with Patricia Vanneste and Simone Kühn. fMRI analyses were done jointly with Simone Kühn.

Abstract

Action decisions can be directly driven by the current state of the external environment (instructed decisions); or they can be driven by internal mental states and goals, independently of the current environment (intentional decisions).

Neural, behavioural and subjective data suggests that two separate neural systems may drive instructed and intentional action respectively. The distinction can be generalized beyond action itself to action *inhibition*, also classifiable as either instructed or intentional.

However, the validity of the instructed/intentional distinction remains controversial. This thesis presents three linked sets of experiments that explored the validity and generality of the distinction, extending it in two key directions: action inhibition and subjective experience. The first group of experiments concerned decision making processes related to action and inhibition; the second focused on the period between decision making and action execution and the third on the subjective experience of intentional actions.

Decision making processes were addressed by comparing electrophysiological and subjective measures prior to, during and after decisions to act or inhibit action. In the absence of external imperatives, intentional decisions may capitalize on spontaneous neural fluctuations, and show a weaker neural code than their instructed counterparts.

To further explore the relative strength of intentional decisions, the period between decision making and action execution was addressed with EEG and behavioural methods. No evidence for unstable intentional decisions was found, during decision maintenance. Intentional action decisions may be strong, and persistent.

Finally, two experiments directly compared the subjective experience of intentional and instructed actions. Neuroimaging results revealed possible mechanisms associated with the subjective experience of acting intentionally.

Together, results support the broad distinction between instructed and intentional decisions. In particular, the coding of intentional actions may involve partial activations of alternative responses. Importantly, the thesis also demonstrates the feasibility of experimental studies addressing the subjective experience of intentional behaviour.

Table of Contents

Chapter 1 General introduction

"Desire, wish, will, are states of mind which everyone knows, and which no definition can make plainer. We desire to feel, to have, to do, all sorts of things which at the moment are not felt, had, or done. If with the desire there goes a sense that attainment is not possible, we simply wish; but if we believe that the end is in our power, we will that the desired feeling, having, or doing shall be real; and real it presently becomes, either immediately upon the willing or after certain preliminaries have been fulfilled.

The only ends which follow immediately upon our willing seem to be movements of our own bodies. …" (James, 1890)

In the general introduction, the literature on neural systems for action and action inhibition is reviewed, focusing on conceptual and functional distinctions between instructed and intentional systems for action. Action inhibition systems can also be classified into instructed or intentional. The roles of subjective reports and neurophysiological data in studying the distinction between instructed and intentional behaviour are considered.

1. 1 Definitions and scope

An action such as moving one's hand may either be a direct and immediate result of an imperative stimulus (e.g., reaching towards a falling object), or may occur for reasons that seem unrelated to any single identifiable stimulus at all, but are instead strongly related to the internal states of the individual (e.g., reaching towards a book that one wants to read).

Motor behaviour can thus be broadly divided in two main categories, namely stimulus driven or *instructed*; and stimulus independent or *intentional*. Instructed behaviour can be directly related to an imperative stimulus. Conversely, intentional behaviour does not obviously relate to any external stimulus but instead to internal states (i.e., desires, intentions). The intentional/instructed distinction recalls the accepted distinction between respondent and operant behaviour. Skinner (Skinner, 1938) first noted that the causes of an animal's behaviour may have come from one of two main sources. First, causal factors of behaviour may be found in the immediate external environment (e.g., food presented causes an animal to approach a person). Alternatively, it may be necessary to consider the internal states of an animal to explain behaviour (e.g., an animal approaches a person in order to be petted). In the first case, a change in the environment directly causes behaviour. In the second case, a change in the animal's internal states is responsible for the observed behaviour.

These internal states are also ultimately related to the external environment, and could thus be seen as representations that mediate between the external world and the expression of behaviour. This mediation means that intentional behaviour can be remote in time from many of the factors that are relevant to its causation, and thus shows "freedom from immediacy" (Shadlen & Gold, 2004).

To address mechanism of motor behaviour, an established experimental tradition has studied action processes. Mechanisms of intentional action processes have been contrasted with those of instructed action, in order to isolate the neural mechanisms underlying volition.

Intentional behaviour requires impulse control

Importantly however, voluntary behaviour very often requires self-control. It is clear that many times it is necessary *not* to respond to stimuli that are encountered in the external world. In other words, in order to be intentional creatures, individuals need to be able to *stop* being externally-driven creatures. For this reason, actions made, or those *not* made, are the result of a balance between positive and negative motor inhibitory components.

Inhibition of action produces, by definition, no behavioural effect. Thus, perhaps for methodological reasons, empirical studies have focussed on intentional action and neglected processes of intentional inhibition. Importantly however, observations from both healthy and clinical populations stress the importance of understanding the mechanisms of intentional inhibition.

Introspection suggests that restraining our impulses can sometimes be very hard. Containing an angry reaction or controlling addictive behaviour is normally effortful. Within the neuropsychological data, some puzzling disorders (such as anarchic hand syndrome and utilization behaviour) reveal deficiencies in action inhibition mechanisms. Therefore, an empirical approach to intentional behaviour should address mechanisms of action inhibition as well as mechanism of action execution.

To approach the study of intentional behaviour, this introduction makes use of three conceptual distinctions. First, a conceptual distinction between *intentional* and *instructed* behaviour is presented. The evidence supporting this conceptual distinction relies mainly on the study of action mechanisms. Therefore the neuroanatomical properties of action systems and their functional relevance are discussed.

Second, a unifying conceptual framework is considered for the study of behaviour (the "what, when, whether" model of action decisions).

Third, a parallel is drawn between *action* and *inhibition* processes. In particular, it is suggested that the distinction between intentional and instructed *action* systems may be mirrored by a distinction between intentional and instructed *inhibition* systems. Consequently, the neuroanatomical properties of inhibition systems and their functional relevance are discussed.

1. 2 Voluntary action

1. 2. 1 Neuroanatomy of action systems

The neuroanatomy of action systems has been extensively studied, and most of the relevant information comes from primate studies. The functional correspondence between human and primate anatomy has been addressed primarily from human neuroimaging studies, but is yet to be fully elucidated.

If a neuroanatomical structure is related to intentional behaviour, it should show two main characteristics. First, it should clearly be related to movement, or inhibition of movement. Second and importantly, it should be related to intentions. That is, its activity should bear a relationship with some higher-level feature of movement, such as target acquisition, goal representation, prospective memory for action, etc. For example, the muscles and the motoneurons that excite them are clearly related to movement execution, but not to the expression of intentions. Therefore, although limb muscles are clearly fundamental for the physical expression of intentions, their electrical activity will say little about how *intentional* control is achieved. Therefore, these two aspects are important in the exploration of volitional systems. Consequently, the relationships between neural activity and both movement and intention are considered in the following description of the motor system.

The main cortical areas identified as directly involved in action systems are the primary motor cortex (M1), the premotor cortex (PMC) and the supplementary motor cortices, including the supplementary and presupplementary motor areas (SMA and preSMA) and the anterior cingulate cortex (ACC) (see figure 1.1).

Subcortical components include the thalamus, striatum, Globus Pallidus, Substantia Nigra and Subthalamic nucleus (see figure 1.2).

All motor output ultimately follows the "final common pathway" (Sherrington, 1906) of α-motoneurons that drive muscle contraction. The cortical neural structures close to the "output" end of the motor pathway are relatively easy to identify, and comprise the primary motor cortex. Neural structures that lie at the "input" end of the motor pathway modulate, drive and organize motor behaviour in a more hierarchical way. These structures are involved in other cognitive functions, and are not as easy to identify. Therefore, a useful approach is to follow the motor pathway in a direction that goes against the flow of information, (i.e. tracking back from the muscle to the source of the control signals), to describe the central contributions to motor control. What therefore follows is a description of the cortical and subcortical areas involved in controlling action and their functional relevance and interconnectivity. The subcortical contribution to action control will then be reviewed, in relation to action as well as to inhibition of action.

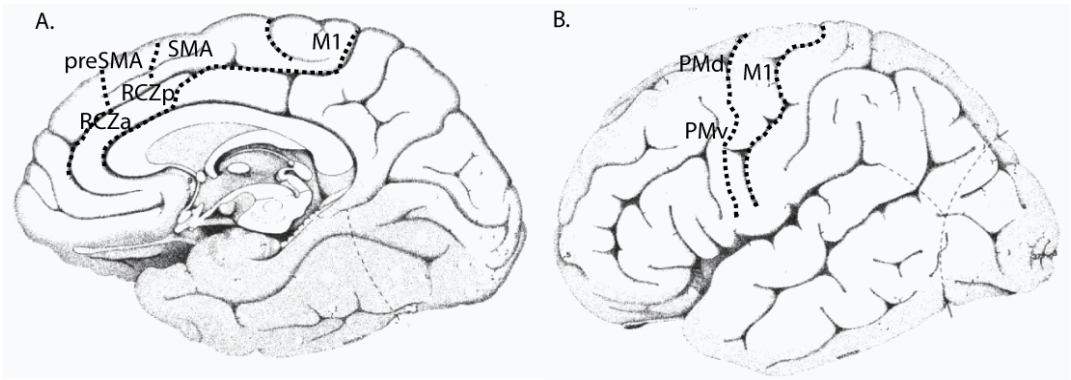


Fig 1.1: A. Medial (right hemisphere) and **B.** lateral surfaces (left hemisphere) showing cortical motor primary and supplementary areas. Primary motor cortex (M1); premotor (PM; dorsal, PMd and ventral, PMv); (pre)supplementary motor area (preSMA); rostro cingulate zone (RCZ; anterior, RCZa and posterior RCZp). RCZ, comprise the cingulate motor areas (CMA). (Adapted from Duvernoy, 1991)

1. 2. 1. 1 Primary motor cortex

The main cortical area that send projections to the spinal cord is the primary motor cortex (M1) (Fulton & Jacobsen, 1935) (See Fig 1.1). M1 possesses direct connections to the spinal cord, premotor areas, parietal cortices, thalamus and cerebellum (Guye *et al*., 2003), and is considered the main source of voluntary cortical motor commands to the body. Penfield and Boldrey (Penfield & Boldrey, 1937) famously mapped the human M1 by direct electrical stimulation in intraoperative settings. Their work showed positive motor signs evoked in specific muscles, leading to the notion that M1 contains a precise somatotopical map of the body (as illustrated by the famous motor homunculus). In line with these findings, M1 was first thought to encode only relatively low-level movement features, and its activity was thought to be modulated only by the muscles required for the action. More recently however, the idea of a strict somatotopy in M1 has been challenged. Although the cortical representation of each muscle is partially segregated in M1, there is also a significant amount of overlap. This has been shown by single-cell recordings in the monkey M1 (Schieber & Hibbard, 1993) and confirmed by human neuroimaging studies, (e.g., Beisteiner *et al*., 2001; Hluštík, Solodkin, Gullapalli, Noll, & Small, 2001). The notion of a strict somatotopy has therefore been replaced by a notion of a certain degree of mosaicism. In other words, brain areas representing specific body areas are not fully segregated but can be found interdigitated. In addition, M1 does not only contain the representation of individual muscles. Instead, the activity of single M1 neurons has been shown to encode simple movements (Kakei, Hoffman, & Strick, 1999). The firing patterns of these "movement" encoding neurons did not depend on the specific muscle activity required for the movement, but on the final target position, in an extrinsic space. Together, these results suggest that M1 may code for some high-level aspects of action, rather than simply the specific pattern of muscle activity required.

Upstream of M1, the premotor areas contribute to drive motor output. They may do so either through direct projections to the spinal cord or indirectly, through connections to M1 (Dum & Strick, 1991). Two sets of neural structures can be identified, and have a functional relevance for the distinction between intentional and instructed behaviour. This distinction will be discussed at length in section 1.2.2. Briefly, lateral premotor areas have been associated with instructed action, and

medial premotor areas with intentional action. The anatomical and functional features of these premotor and supplementary motor areas will be discussed in what follows.

1. 2. 1. 2 Lateral premotor areas: premotor ventral and premotor dorsal surfaces

The lateral premotor areas lie immediately upstream of M1. Consistent with monkey neuroanatomy (Dum & Strick, 1991), in humans the lateral surface of each hemisphere contains the premotor ventral (PMv) and premotor dorsal (PMd) surfaces. Tomassini *et al* (Tomassini *et al*., 2007) have studied the functional connectivity of the dorsal and ventral PMC through diffusion-weighted imaging tractography. Their results show that the connectivity profile of PMd is different from that of PMv. The PMd is most strongly connected to superior parietal areas and dorsal prefrontal areas, whereas the PMv is mainly connected to inferior parietal lobule and ventrolateral and orbital prefrontal areas. The connectivity-based distinction is consistent with functionally-derived distinctions (Mayka, Corcos, Leurgans, & Vaillancourt, 2006).

Several monkey studies support a functional distinction between PMd and PMv. It has been suggested that the functions of PMv and PMd differ in terms of their integrative role in sensorimotor processing (Hoshi & Tanji, 2007). On the basis of monkey single-cell electrophysiology, Hoshi and Tanji suggest that the PMv has a simple or "direct" integrative function, and sends motor outputs to drive an action that directly matches a simple source of information. By contrast, they suggest that PMd has a major role in "indirect" sensorimotor processing, retrieving multiple sets of motor information from sensory signals, and integrating components of a required action to formulate a motor program for the intended action.

In humans, the functional distinctions between PMd and PMv are less clear. One transcranial magnetic stimulation (TMS) study has yielded results consistent with the division of labour suggested by the monkey data (Davare, Andres, Cosnard, Thonnard, & Olivier, 2006). Transient virtual lesions (caused by repetitive TMS) of

PMv impaired specifically precision grasping functions, whereas repetitive TMS over PMd impaired the timing and coordination of different subcomponents of the movement. This result is consistent with a relatively more integrative role for PMd, and a lower-lever movement control role for PMv. However, both PMv and PMd in humans have been related to higher level motor functions, such as action imagery and mental hand rotation (Abe & Hanakawa, 2009; Rizzolatti, Fogassi, & Gallese, 2002).

In addition to the premotor areas of the lateral surface, three structures within the medial surface have been related to motor control. These are the supplementary motor area (SMA), located caudal to the anterior commissure; the preSMA, rostral to the anterior commissure and the cingulate motor areas (CMA).

1. 2. 1. 3 Medial premotor areas: SMA and preSMA

The first two medial areas that lie immediately upstream of M1 are the supplementary motor areas. The key difference between SMA and preSMA is their connectivity patterns (Picard & Strick, 2001). In humans, diffusion weighted imaging has shown that SMA and preSMA present clearly different connectivity profiles (Johansen-Berg *et al*., 2004). SMA was found to project to the spinal cord and M1, whereas only preSMA, showed connections with the prefrontal cortex and the medial parietal cortex. Both SMA and preSMA presented connections with the thalamus, but preSMA is connected to more anterior parts. These different properties vary along a rostrocaudal gradient, rather than showing a sharp separation between the two areas.

Results from human neuroimaging studies consistently suggest that SMA and preSMA also differ in their functional properties. Whereas increases in blood-oxygen level dependent (BOLD) activity in the SMA depend on aspects of movement behaviour e.g. movement force (Dettmers *et al*., 1995); preSMA activations are normally associated with the cognitive aspects of the tasks (Jenkins, Jahanshahi, Jueptner, Passingham, & Brooks, 2000).

A final difference between SMA and preSMA is their degree of somatotopical organization. In humans, evidence for a coarse somatotopy comes from both direct electrical stimulation (Fried *et al*., 1991) and neuroimaging studies of movement of hand and foot (Cauda, Geminiani, D'Agata, Duca, & Katiuscia, 2011). Interestingly, the neuroimaging study by Cauda *et al* showed that BOLD activity increases in ventral M1 (hand area) are accompanied by BOLD activity increases in rostral SMA. More dorsal M1 (foot area) BOLD activity levels are in turn accompanied by more caudal SMA activity, suggesting a somatotopical coupling between SMA and M1. In contrast there are no clear reports of somatotopy or effector-specificity in preSMA.

In sum, there is considerable evidence suggesting that SMA is related to lower-level aspects of movement behaviour such as movement parameters, whilst preSMA seems to play a role in diverse higher-level cognitive tasks, such as action planning.

The precise role of the preSMA is still a matter of debate (Nachev & Husain, 2010; R. E. Passingham, Bengtsson, & Lau, 2009). Three different functions have been proposed for the preSMA. Some of these seem to be simply different descriptions of the same function, so they can be grouped in three broad categories. These are in general terms (1) intention representation, (2) action plan maintenance and (3) response inhibition.

First, increased preSMA BOLD activity has been associated with action intention representation, in cases of both the representation of our own intentions (H. Lau, Rogers, Haggard, & Passingham, 2004), action observation (Cunnington, Windischberger, & Moser, 2005) and voluntary action selection (Soon, Brass, Heinze, & Haynes, 2008; Zhang, Hughes, & Rowe, 2012). Intriguingly, Fried *et al* (Fried *et al*., 1991) found that direct electrical stimulation of SMA could induce conscious "urges" to move, suggesting a link between the SMA and the generation or introspection of intentions.

Second, preSMA activity in either monkeys and humans has been related to motor learning (Hikosaka, Nakamura, Sakai, & Nakahara, 2002), maintenance in memory of action plans (Stadler *et al*., 2011), visuo-motor associations (Sakai et al., 1999), action sequence initiation (Kennerley, Sakai, & Rushworth, 2004) and updating of motion plans (Shima, Mushiake, Saito, & Tanji, 1996).

Third, preSMA has been related to response inhibition (Aron & Poldrack, 2006) and, more generally, to conflict resolution (Nachev, Kennard, & Husain, 2008). It has therefore been suggested that the diverse functions associated with increased preSMA BOLD activity should be parsimoniously interpreted as conflict resolution (Nachev, Rees, Parton, Kennard, & Husain, 2005).

The disagreement in the precise function attributed to the preSMA may depend on the specific task demands, as several different functions may be subserved by one given area. Whatever function is associated with preSMA, it is clear that it plays an important role in volitional control in situations of ambiguity or the need for response selection, initiation or control.

1. 2. 1. 4 Cingulate motor areas

The cingulate motor area (CMA, see figure 1.1) lies anterior to the supplementary motor cortices. The CMA lies within the anterior cingulate cortex and projects to supplementary motor and premotor cortices as well with M1 and the spinal cord (Picard & Strick, 1996). This connectivity profile strongly suggests that the CMA is involved in motor control. In humans, CMA can be subdivided in three different zones that differ on their cytoarchitectonic properties and function (Vogt, Nimchinsky, Vogt, & Hof, 1995). Picard and Strick have distinguished the caudal cingulate zone (CCZ, behind the VCA line) from the rostral cingulate zone (RCZ), in turn subdivided into its anterior and posterior segments (RZCa and RCZp respectively).

CCZ and RCZ differ in their connectivity patterns. In a recent resting state functional connectivity analysis, Habas (Habas, 2010) has shown that CCZ and RCZ share a large common connectivity pattern, including connections with limbic and sensorimotor regions. However, RCZ showed more widespread connections with prefrontal, premotor and parietal cortices, whereas CCZ presented more widespread connections with sensorimotor cortex. This connectivity pattern is consistent with results from functional studies. The CCZ has been mainly related to movement execution (Picard & Strick, 2001); whereas the RCZ has been associated with conflict resolution and action selection (Mueller, Brass, Waszak, & Prinz, 2007;

Picard & Strick, 2001). Further, it has been suggested that RCZa is involved in conflict monitoring, and the posterior portion playing a role in action selection (Picard & Strick, 2001). A recent study has extended Picard and Stick's (Picard & Strick, 2001) original meta-analysis and has provided further support for this distinction (Beckmann, Johansen-Berg, & Rushworth, 2009).

Both RCZa and RCZp present some coarse somatotopy and BOLD activity increases corresponding to arm or face movement that can be spatially segregrated in both anterior and posterior RCZ (Picard & Strick, 1996; Shackman *et al*., 2011).

1. 2. 1. 5 Prefrontal Cortex

A final structure clearly involved in motor control is the prefrontal cortex. Although it is not strictly considered a motor area, motor and premotor areas are affected by inputs from the prefrontal cortex (PFC) (Goldman-Rakic, 2007).

Areas within the PFC may be taken to occupy the highest levels in the hierarchy of motor planning. It is important to point out that rather than being "homuncular" in nature, the PFC has mainly an associative function and receives input from limbic, sensory and associative cortices.

PFC function has been associated with most aspects of behavioural guidance, such as processing and integration of both directly perceived and memorized information, associative learning, reward based behavioural control and decision making, (see Tanji & Hoshi, 2001) for a review). Neuronal activity in PFC has been shown to be related to coordination of concurrent information, allowing for a temporal sequencing of multiple actions and goal and action selection (Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Tanji & Hoshi, 2001). Evidence from human neuroimaging data is consistent with that from monkey electrophysiology. For example, it has been proposed that PFC plays a role in achieving goals by maintaining behavioural goals in working memory and representing the actions necessary to reach those goals (Miller & Cohen, 2001).

It has been argued that the PFC allows for response selection by "sculpting the response space" (C. Frith, Gallagher, & Maguire, 2004). Several pieces of evidence

support this notion. In one fMRI study, Desmond et al (Desmond, Gabrieli, & Glover, 1998) tested participants in a word completion task. They found that DLPFC showed greater levels of BOLD activity with larger numbers of possible response alternatives, or larger response spaces. Further, in one positron emission tomography (PET) study participants were tested in a sentence completion task (Nathaniel-James & Frith, 2002). The presented sentences varied in their degree of constraint. I.e., some of the sentences presented strongly suggested only one possible response alternative, whereas others admitted several response alternatives. Together, these results suggest that DLPFC is involved in response selection, with increased effort necessary for the "pruning" of greater response spaces.

In addition, Koechlin *et al* studied the organization of the information processing in PFC (Koechlin, Ody, & Kouneiher, 2003). The authors asked participants to make responses to external stimuli. Crucially, in separate blocks, the responses required an increasing level of complexity in the processing of the necessary response. The results show that higher processing complexity was associated with more frontal increases in BOLD activity within the PFC. On the bases of these results, the authors suggest a hierarchical model for PFC function (Koechlin & Summerfield, 2007).

1. 2. 1. 6 Subcortical structures: Basal ganglia

The basal ganglia (BG) are located at the base of the frontal lobes and contribute to motor control, sending feedback to the cortex and participating in coordination and inhibitory functions. They consist of four main subnuclei: striatum, globus pallidus (in turn comprised of the internal segment -GPi- and external segment -GPe-), subthalamic nucleus (STN), and substantia nigra -including the *pars compacta* (SNc) and *pars reticulata* (SNr)- (See Fig 2).

The primary and premotor cortices (M1, SMA, PM) and the somatosensory cortex project excitatory outputs onto the striatum. The striatum, in turn, sends inhibitory projections to two output nuclei, the internal segment of the globus pallidus (GPi) and the substantia nigra pars reticulata (SNr), through two possible parallel pathways, known as "direct" and "indirect". The direct pathway consists of monosynaptic connections from the putamen to the GPi/SNr; whereas the indirect

pathway consists of polysynaptic connections involving the external segment of the globus pallidus (GPe) and the subtalamic nucleus (STN). Inhibitory projections from the GPi/SNr then reach the thalamus, which in turn projects its excitatory output back into the cortex (Spence, 2009).

A widely accepted view includes a third possible connection between the cortex and the STN, i.e. the hyperdirect pathway (Aron & Poldrack, 2006; Chambers, Garavan, & Bellgrove, 2009). This pathway is a monosynaptic excitatory connection from the cortex to the STN, bypassing the striatum. The hyperdirect pathway has special relevance for action inhibition (see section 1.3).

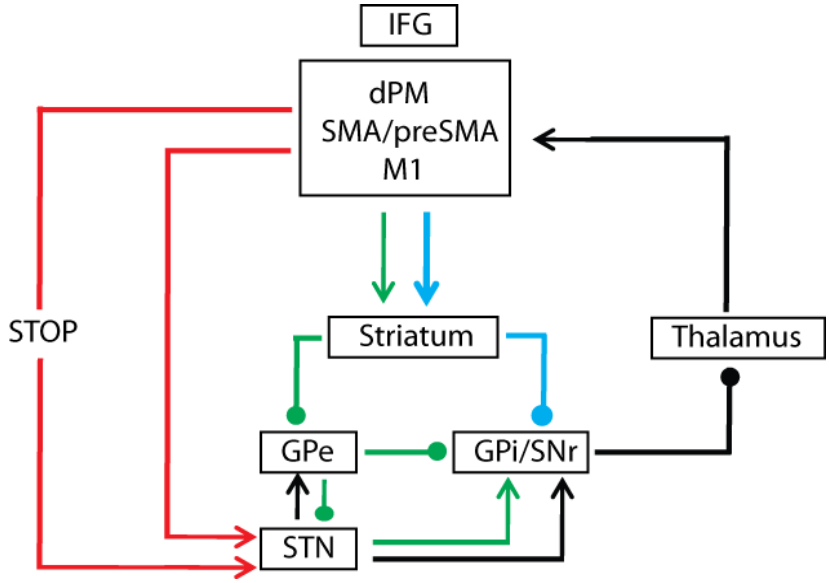


Figure 1.2: Box diagram indicating basal ganglia connectivity. Arrow connectors indicate excitatory connections. Dot connectors correspond to inhibitory connections. Red, blue and green lines correspond to the hyperdirect, direct and indirect pathways respectively. Adapted from Chambers *et al*., (Chambers *et al*., 2009). Represented subcortical structures include subthalamic nucleus (STN); globus pallidus (GP internal -GPi- and external -GPe- segments); subtantia nigra *pars reticulata*.

1. 2. 2 Principles of action control

Given the neuroanatomical systems that support motor control, in this section a general framework for the investigation of intentional action is provided.

Every action decision includes at least three dimensions, what action to make (*what* component) when to make it (*when* component) and whether to make the action at all (*whether* component) (Brass & Haggard, 2008). These three dimensions are described in the context of the "what, when, whether" (WWW) model for action selection (see figure 1.3).

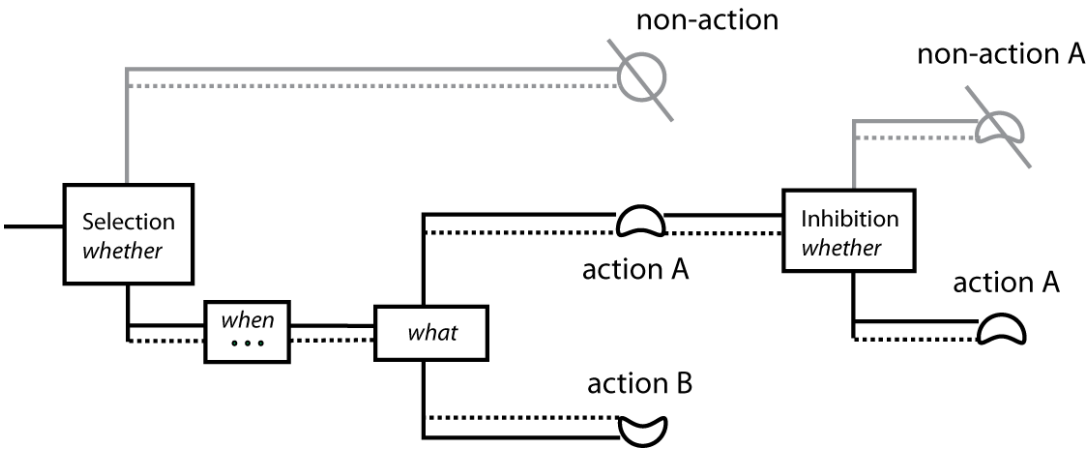


Figure 1.3. Decision tree for action and inhibition, in the context of the "what, when, whether" (WWW) model for action selection. Black sections correspond to actions and grey sections correspond to inhibition of actions. Early *whether* decisions to act or not and *what*, *when* and late *whether* decisions are shown. "*Selection whether"* and "*inhibition whether"* decisions are two separate processes, with the former involving response selection and the latter requiring action inhibition (see text for details). *What* and *when* decisions need not be sequential in time and the precise order may depend on the task. Each decision can be reached by two possible extreme pathways, namely instructed or intentionally. Solid lines represent instructed decisions, dotted lines represent intentional decisions. See section 1.3.4 for example implementations of this decision tree.

Supporting a conceptual distinction, different action decisions have been associated with different brain structures (see figure 1.4). Hoffstaedter *et al* (Hoffstaedter, Grefkes, Zilles, & Eickhoff, 2012) have used fMRI to identify brain areas

differentially involved in the selection of either the *what* or *when* components. Their results show greatly overlapping, but also partially segregated brain structures associated with each decision. The ACC, bilateral IPL and DLPFC are involved in the selection of both the *what* and *when* components. Hoffstaedter *et al* also show that pre-SMA is specifically associated with the *what* component; whereas internal timing (*when* component) relies crucially on bilateral anterior putamen and globus pallidus and on a well-distributed timing network comprised of bilateral area 44 and anterior insula for cognitive time processing and SMA, basal ganglia, and cerebellum related to more automated timing of movement execution.

To formally examine a dissociation between the *what* and *when* components of action, Krieghoff et al devised a task in which each of these components could be manipulated independently (Krieghoff, Brass, Prinz, & Waszak, 2009). In each trial, two visual cues could either specify (instructed conditions) or let participants choose (intentional conditions) the *what* and *when* components. This resulted in four possible combinatorial conditions. The authors contrasted BOLD activations between intentional and instructed conditions for each component independently, at the time of presentation of the visual cue. They found increased BOLD activity in RCZ when participants selected *when* to act. Conversely, the authors found increased BOLD activity in the superior frontal gyrus (SFG) in the paramedian frontal cortex, when participants decided *what* action to make.

To investigate a distinction between *what* and *whether* decisions, Kühn et al (Kühn, Haggard, & Brass, 2009) did an fMRI experiment in which participants could intentionally decide whether to press a key or inhibit the key press. They found that the dorsal frontomedian cortex (dFMC) showed increased levels of BOLD activity in trials in which participants decided to inhibit a key press. This suggests that dFMC activity is related to the implementation of inhibition of action. In addition, increased RCZ BOLD activity was associated with decisions to act and to inhibit, as compared with the instructed cases. The authors argue that RCZ is involved in the response selection process, regardless of the outcome. This study will be further discussed below for its implications for the understanding of mechanisms for action inhibition (see section 1.3.4.3)

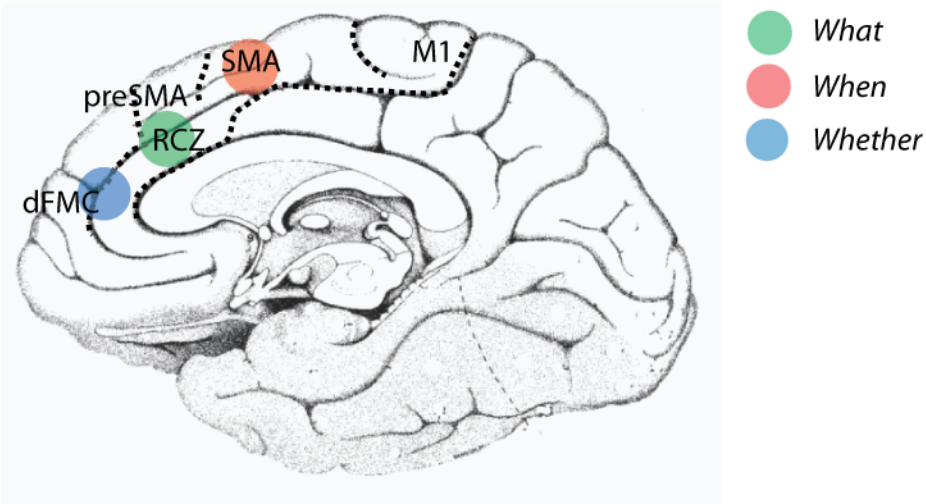


Figure 1.4 Medial surface of the brain(right hemisphere) showing the neuroanatomical structures associated with intentional *what* (rostro cingulate zone, RCZ) *when* (supplementary motor area, SMA) and late *whether* (dorsal fronto median cortex, dFMC) decisions. Adapted from (Duvernoy, 1991)

Importantly, the *whether* component of action decision can occur at two rather different stages of action preparation, suggesting that there are two possible sources of non-action: (1) *early decisions* about whether or not to begin action processing ("selection *whether*") and (2) *late decisions* about whether to inhibit a final motor output ("inhibition *whether*"). The former occurs before any action preparation takes place and can be explained in terms of action selection processes alone, while the latter requires an additional process of intentional inhibition. Late decisions to inhibit would have the specific function of blocking motor output, and suppressing an action that has already been prepared.

Regarding early *whether* decisions, neuroimaging evidence suggests that a decision to omit an action shares some brain correlates with decisions to act. In this way, decisions to "do nothing" may involve active action selection processes, but not necessarily inhibition components. Kühn and Brass (Kühn & Brass, 2009) asked participants to make decisions about whether to act or not upon the appearance of a "decide" signal. These decide signals were presented in the context of a task with a majority of simple reaction trials. Therefore the decide signals required that participants stopped their prepotent impulse to act, and subsequently made an early *whether* decision, to select between acting or not. Kühn and Brass measured the

BOLD activity associated with these early *whether* decisions, and found that early decisions to act (go) only differed from early decisions not to act (nogo) in their activation of areas related to motor execution, strongly suggesting that both go and nogo decisions included action selection processes. The task included an additional stop signal that required participants to stop their reaction, but did not require a subsequent late *whether* decision. The authors then contrasted the BOLD activity associated with trials involving early whether decisions with that of trials involving only simple stopping. A conjunction analysis revealed that early *whether* decisions (regardless of the outcome) recruited bilateral insular cortex, right DLPFC and RCZ. These areas are typically associated with action selection tasks. These results suggest that decisions not to act share the decision processes with decisions to act, and nonaction should therefore be considered as an alternative action outcome.

Two further studies support this idea. First, Kühn *et al* (Kühn, Bodammer, & Brass, 2010) have shown that multivariate pattern analysis classification algorithms can distinguish the pattern of BOLD activity associated with active decisions not to act from that associated with not deciding at all. Predictive areas included SMA, precuneus and right inferior frontal gyrus (rIFG). These areas had not been identified by a classical univariate analysis comparing trials including decisions not to act with those including no decisions at all, indicating that the classification power did not depend simply on generalized BOLD activity increases in broad brain regions.

Second, in a series of experiments, Kühn *et al* have shown that nonactions can be bound to perceptual effects that have been arbitrarily associated with them, showing the same kind of action-effect binding demonstrated by Elsner and Hommel (Elsner & Hommel, 2001). Elsner and Hommel induced stimulus-response associations (S-R) by asking participants to freely choose to make one of two actions, and presenting tones of different pitches contingent on the action selected. In a subsequent test phase, participants were faster and more accurate in making actions in response to tones with a S-R consistent with the one learnt during the acquisition phase as compared to when the required S-R was inconsistent with the one of the acquisition phase. In the same way, Kühn *et al* (Kühn, Elsner, Prinz, & Brass, 2009) have shown that nonactions show a "nonaction-effect" binding. Importantly, this binding is dependent on participants forming an intention not to act, as instructed nonactions

require no active *whether* decisions and generate no binding. Further, nonactions seem to have specific representations, as "not-moving a right hand" may be associated to a different effect than "not-moving my left hand" (Kühn & Brass, 2010). Interestingly, these specific non-action representations seem to be possible even at the stage of an early *whether* decision. This suggests that action preparation may not be a necessary requirement for non-action effect binding.

In contrast to early *whether* decisions, that are related to action selection, late *whether* decisions are tightly related to action inhibition and will be discussed in section 1.3.4.

1. 2. 2. 1 Experimental operationalizations of intentional and instructed actions

An established empirical tradition has drawn objective distinctions between intentional and instructed action. In this tradition, instructed actions are operationally defined as arbitrary but fixed associations of e.g. visual stimuli to simple movements. Typically, in instructed actions, both the timing and type of action are explicitly specified (the *when* and *whether*, figure 1.3). In contrast, intentional actions are usually operationalized by leaving one of the dimensions underspecified (*when*: e.g. Libet, Gleason, Wright, & Pearl, 1983, Thaler, Chen, Nixon, Stern, & Passingham, 1995; *what*, e.g. Lau, Rogers, Ramnani, & Passingham, 2004; Mueller et al., 2007, *whether*, e.g., Kühn, Haggard, et al., 2009). The participants must internally decide whether to move, what movement to make, or when to make it, and must then generate the movement on the basis of this internal information. As such, by definition intentional actions are not fully specified by any single environmental event.

Within this operational framework and supported by evidence from functional, neuroanatomical, cytoarchitectonic and ontogenic features of the primate brain, Goldberg (Goldberg, 1985) proposed the existence of two parallel systems for action. He suggested that the lateral premotor cortices (PMd and PMv) were responsible for instructed or reactive actions, i.e., those that are driven by explicit external signals. Goldberg further suggested that medial premotor areas, notably the SMA, were responsible for intentional, self-generated or "projectional" action.

Whilst a sharp division between lateral and medial aspects is clearly artificial, this conceptual framework is a useful guide to approach a description of the neuroanatomical systems involved in motor control.

In the context of the WWW model for action control discussed above, instructed actions will be those that fully follow specified instructions, constraining all three action dimensions (*what*, *when*, *whether*). In contrast, intentional actions will be generated and made, despite some ambiguity in one or several of the dimensions of action decisions. For example, a low-pitch tone may signal that a right finger movement should be made at the time of the tone. In this case, all three components, namely *what* (right and not left finger), *when* (at the time of the tone) and *whether* (yes) dimensions are specified by the external auditory instruction. In contrast, a low-pitch tone may indicate that a movement should be made with *either* finger, thus leaving the *what* dimension unspecified. Similar comparisons can be made for the *when* and *whether* dimensions.

1. 2. 2. 2 Experimental evidence for a distinction between intentional and instructed actions

Several lines of research provide evidence supporting Goldberg's hypothesis and will be described in this section. In monkeys, robust evidence from lesion studies supports a functional distinction between instructed and intentional systems. In humans, evidence for a distinction comes from behavioural measures, neuroimaging data and electrophysiological (both scalp and intracranial) recordings. This converging evidence will be described to provide support for a conceptual distinction between instructed and intentional action.

In a set of pioneering studies, Thaler *et al* (Thaler *et al*., 1995) trained monkeys to make spontaneous, self paced movements in order to get a food reward. Thaler *et al* showed that removal of the medial premotor cortex in monkeys dramatically reduced the rate of such spontaneous movements. In contrast, when monkeys were trained in the same task (to make the same movement and obtain the same food reward), but in response to external auditory cues, medial premotor cortex lesions virtually did not affect the monkeys' spontaneous movement rate. Therefore, the lesion in the medial

premotor cortex did not affect the monkeys' ability to make the movements *per se*, or even to be driven by the food reward. Instead, these results suggest that the lesion impaired the monkeys' ability to generate arm movements intentionally. Passingham *et al* (Passingham, 1987) replicated and extended this result. They trained monkeys to make arm movements in the absence of external visual cues. The monkeys sat in the dark and raised their arms to roughly eye-level in order to get a food reward. Bilateral removal of the supplementary motor cortex affected the monkeys' ability to make accurate movements driven solely by proprioceptive input. When the lights were turned on and monkeys could use visual cues to guide their movements, they performed like unoperated animals.

In addition and importantly, Passingham *et al* (Passingham, 1985) trained other monkeys to perform a visually guided motor task. A blue or red light required monkeys to pull or turn a handle respectively. Monkeys with bilateral premotor lesions could execute the movements, but were unable to do so in response to the visual cues. Again, the lateral premotor lesion did not affect the monkeys' ability to make the movements *per se* but to pair these movements to arbitrary external cues.

In humans, behavioural evidence also supports a distinction between instructed and intentional action systems. Obhi and Haggard (Obhi & Haggard, 2004) asked participants to make self-paced key presses. Occasionally, they interrupted these self-paced intentional actions with an instruction that required a switch from an intentional action mode to an instructed action mode, but not a change in the movement effectors. The authors found a reaction time (RT) cost of this switch that they coined "truncation effect". This suggests that instructed action systems cannot benefit from existent action preparation of intentional action systems. However, this result has recently been challenged (Hughes, Schütz-Bosbach, & Waszak, 2011) by the results of a similar experiment. Hughes *et al* found a truncation effect only if the switch between intentional and instructed action plans also involved a switch between movement effectors, suggesting that instructed actions may indeed benefit from processes of preparation of intentional actions, and questioning the strong dissociations drawn between the two systems.

1. 2. 2. 3 Neuroimaging evidence for a distinction between instructed and intentional action

Several human neuroimaging studies have addressed the neural basis of the difference between intentional and instructed actions, and the results have mirrored those from monkey lesion studies. The contrast between intentional *vs*. instructed action has been associated most consistently with increased BOLD signal in SMA, preSMA and RCZ (Cunnington, Windischberger, Deecke, & Moser, 2002; H. Lau, Rogers, Ramnani, et al., 2004; H. Lau, Rogers, & Passingham, 2006a); see (Krieghoff, Waszak, Prinz, & Brass, 2011) for a review. In particular, Müller *et al*. (Mueller *et al*., 2007) have suggested that RCZ is mainly involved in selecting the *what* component; whilst preSMA participates in the selection of the *when* component (see figure 1.4).

The inverse contrast, of instructed vs. intentional actions has shown less consistent patterns of activity. Debaere (Debaere, Wenderoth, Sunaert, Van Hecke, & Swinnen, 2003) found increased BOLD signal in dorsal premotor cortex, in line with Goldberg's hypothesis; but this result has not been as widely reported as the increased BOLD signal in medial and parietal aspects of the brain associated with the inverse contrast. Krieghoff *et al*. (Krieghoff *et al*., 2011) speculate that this may be due to the fact that instructed actions are more automatic and require less cognitive effort, and they therefore elicit less of a differential BOLD activity.

In summary, the functional distinction between intentional and instructed action systems is often based on the neuroanatomical distinction between a medial frontal system for intentional action, centred on the SMA, and a lateral premotor system for instructed action, centred on the premotor cortex.

Medio-lateral distinctions related to the amount of information to be integrated are not restricted to motor areas. In the PFC differences between medial and dorsal systems has also been reported. Interestingly, Koechlin *et al* (Koechlin, Corrado, Pietrini, & Grafman, 2000) found that increases in BOLD activity differed in the medial and lateral PFC, according to the nature of the task. The medial PFC showed increased activation with a task that required responses that could be prepared and expected, whereas the lateral PFC showed increased activation with tasks that

required responses that were unpredictable, and contingent on the external environment.

In monkeys, the dorsal and ventral parts of the lateral PFC present different connectivity patterns. The lateral PFC is part of an orbitoventral network that receives multiple sensory inputs, from visual, auditory, somatosensory and olfactory and gustatory modalities. In contrast, the dorsal PFC is part of a mediodorsal network that receives input from multimodal areas in the temporal cortex (Tanji & Hoshi, 2008). This suggests that the dorsal PFC receives signals that have already been through a processing stage, whereas the lateral PFC processes relatively "raw" inputs, and that a medial/lateral distinction may not be confined to the motor structures.

1. 2. 2. 4 Studies of the subjective experience of intentional action

One of the few studies exploring potentially differential effects on subjective experience of intentional action studied the phenomenon of intentional binding. In intentional binding, participants make actions that are paired to a given effect, such as a tone. When participants are asked to retrospectively judge the time of their action and the time of the action effects, the actions are judged later, and the effects earlier, than when they actually happened. In other words, action and effects are bound together in time (Haggard, Clark, & Kalogeras, 2002). When actions are not generated by the participants but artificially, by means of a TMS pulse to M1, the temporal binding effect is reversed. Wenke *et al* (Wenke, Waszak, & Haggard, 2009) asked whether temporal binding effects would distinguish between intentional and instructed actions. Participants did a modified temporal binding task, in which they were either instructed or free to choose the *what* (right/left hand) or *when* (first/second interval) components in each trial. Interestingly, the authors found no differences in terms of intentional binding between intentional and instructed actions, suggesting that intentional binding does not depend on the action selection processes, but rather on motor preparation and effect anticipation.

1. 2. 3 Electrophysiology of action

Scalp electroencephalography (EEG) recordings time-locked to the time of movement reveal the readiness potential (RP), a slow negative potential that peaks around the time of movement (Kornhuber & Deecke, 1965). The RP has been typically related to action preparation and has two main pre-movement segments, the early RP and the late RP (Shibasaki & Hallett, 2006). The early RP typically starts at around 1.5 s before movement onset. It is maximal in fronto-central electrodes and is symmetrically distributed. It shows larger amplitudes and/or earlier onset in conditions of increased level of intention, preparation, movement selection, required forced and learning. The late RP typically starts at around 500 ms before movement. It is characterized by an increase in slope of the RP, and by a lateralization of its topography. The amplitude of the late RP is greater at electrodes that are contralateral to the movement, as compared to ipsilateral electrodes. It is hence referred to as lateralized readiness potential (LRP) and can provide a measure of covert action selection. The RP has been related to increased BOLD activity in preSMA, bilateral SMA, and ACC (Cunnington et al., 2005; Jahanshahi et al., 1995; Jenkins et al., 2000).

RPs for actions following the intentional selection of the *when* component show larger amplitudes than those for actions made at instructed and unpredictable times, it was first taken to be a signature of voluntary action control, e.g. (Jahanshahi et al., 1995). It is now however increasingly clear this is not the case. In an elegant experiment, Baker *et al* (Baker, Piriyapunyaporn, & Cunnington, 2012) asked participants to make two self-paced actions per trial, in order to reproduce a time interval. The experimenters emphasized the need to time the second key press as accurately as possible, in order to reproduce the time interval. No attention was drawn towards the first key press. In this way, the first action started each trial and was somewhat incidental to the task, so it did not require much attention from the participants' part. The second action was kinetically equivalent and also self-paced, but did require attention and precise action preparation. The authors found larger RP amplitudes related to the second action than to the first one, although they were both intentional actions. This strongly suggests that the large RP amplitudes found in

intentional actions may be related not to intentional processes *per se*, but rather to attentional factors that often accompany intention.

Evidence from intracranial electrodes in human patients complements the scalp electrophysiological data. Recently, Rosenberg-Katz *et al* (Rosenberg-Katz *et al*., 2012) have applied a combination of fMRI and intracortical electrophysiological (iEEG) recordings to demonstrate differences at the neural level between instructed and intentional actions. The authors showed that the pre-SMA (but not the SMA) and DLPFC showed higher levels of BOLD activity during intentional compared to instructed action planning. In addition, intentional action planning was associated with higher levels of functional correlation between preSMA and DLPFC, and SMA and DLPFC Finally, iEEG data showed increased inter-regional gamma-related connectivity between electrodes situated in medial and lateral aspects of the prefrontal cortex for intentional compared to instructed actions.

It has been shown that the free selection of the *what* component may also affect RP amplitude (see (Lang, 2003) for a review). Praamstra *et al* (Praamstra, Stegeman, Horstink, Brunia, & Cools, 1995) asked participants to make four types of joystick movements. Participants were asked to make either single or sequential movements. In turn, both single and sequential movements could either be instructed or intentional. That is, participants were either instructed or allowed to freely choose which movements they would make. Praamstra *et al* measured the amplitude of the RPs associated with each one of these movement types. They found higher RP amplitudes for intentional movements as compared to instructed movements. These differences were apparent as early as 1.5 s before the time of movement and were generalized topographically, but concentrated on the centro-parietal electrodes. In addition, single movements showed greater differences between instructed and intentional conditions than did sequential movements. Praamstra *et al* replicated the results when participants were asked to make finger flexions with either hand. These results suggest that the intentional selection of the *what* component enhances RP amplitude. The authors suggest that this may be related to greater involvement of the SMA in actions where the *what* component is intentionally chosen, compared to instructed.

Dirnberger *et al* (Dirnberger, Fickel, Lindinger, Lang, & Jahanshahi, 1998) noted that in experimental paradigms such as the one reported by Praamstra et al (Praamstra et al., 1995), intentional actions are characterized by a non-repetitive sequence of movements. In contrast, instructed actions are made in the context of a repetitive sequence of actions. Consequently, the relatively low RP amplitudes associated with instructed actions may have to do with motor habituation effects rather than with volitional components. To examine this possibility, Dirnberger *et al* recorded EEG activity while participants were making brisk finger movements in order to press one out of four possible buttons. They replicated the finding that RPs show larger amplitudes and earlier onsets in conditions of intentional actions as compared to instructed action. In comparing RPs from trials following long periods of repetitive actions (and consequent motor habituation) with trials that did not follow habituation periods, the authors found that RPs were significantly reduced, suggesting that habituation factors modulate RP amplitude. Critically, the authors found that the LRPs differed strongly between intentional and instructed conditions, with intentional actions showing greater LRP amplitudes. These differences appeared as early as 1.5 s prior to movement.

Together, these results suggest that relatively low-level movement features, such as movement habituation, or attention allocation may affect RP amplitude. This has led to the idea that RPs do not reflect intentional processes. Instead, they may reflect other lower-level attention-related processes that are incidentally associated with intention. Importantly, these low-level features may be central to the "intentional" component of action. For example, intentional actions are intrinsically non-routine, and non-habitual. As such, novel intentional actions will demand more attentional resources than routine actions. The two processes of intention and attention may be difficult to disentangle in the two situations. One fMRI experiment is useful to address this potential concern. Lau et al (H. Lau, Rogers, Haggard, et al., 2004) asked participants to make self-paced actions. The authors reasoned that paying attention to an endogenous cognitive process will increase BOLD activity in the area subserving that process. Thus, in separate blocks, participants were asked to pay attention to the timing of either their action intentions or of the actions themselves. Lau et al found increased SMA activity when participants paid attention to their own intentions, as compared to when they attended to their actions. SMA activity has

been linked to RPs (see above). This elegant result therefore suggests that attentional and intentional processes may be in fact distinguishable, but at the same time intrinsically confounded in several of the classical experiments reviewed here.

1. 2. 4 What is the "intentional" component in intentional action?

Despite the attempts to develop an unambiguous and objective experimental approach to action systems, the nature of the difference between intentional and instructed actions is still debated. Lau and Passingham (R. E. Passingham et al., 2009) have argued that intentional actions are clearly distinct from instructed actions, and that the neural structures that support them are in turn distinct.

Two main criticisms to this argument have been raised. First, it has been argued that given the complexity and obscure nature of the processes underlying "intentional" action, this label is not an appropriate one. Because saying that an action is intentional does not describe in any detail the true processes required for it, the term adds no explanatory power. Roepstoff and Frith (Roepstorff & Frith, 2004), first pointed out that when a participant is asked to make a series of "intentional" actions, she is in fact asked to unpack the obscure instruction to "act as if she had free will". Folk knowledge indicates that apparently "free" action should also be apparently "random" (Ebert & Wegner, 2011). Therefore, the neural mechanisms identified with intentional action defined in such a way may therefore have more to do with the conscious and carefully planned simulation of seemingly random behaviour rather than with true generation of behaviour in underdetermined conditions (Jahanshahi, Dirnberger, Fuller, & Frith, 2000). In line with this view, Nachev and Husain (Nachev, 2010; Nachev & Husain, 2010) argue that in cases of intentional action, what is in fact covertly taking place is a rich series of integrative processes and decisions. Further, they hold that what is particularly unfortunate from an empirical point of view is that these covert integrative processes are scientifically intractable and not accessible to controlled manipulation.

A second criticism points out that a stark distinction between intentional and instructed action is artificial. Even if the two action systems exist as such, they will tightly interact and contribute to drive any given action. There is some consensus on

the fact that intentional and instructed actions are two extremes of a continuum, but it is not clear along which dimension this continuum should be defined. Passingham *et al* agree (Passingham, Bengtsson, & Lau, 2010) that actions may gradually vary in terms of the relative contributions of "exogenous" (environmental i.e., available as simple cues in the perceptual environment) and "endogenous" (internal cues i.e., memory, goals, etc). In fact, Thut *et al* (Thut et al., 2000) have shown that EEG activity between intentional and instructed actions differs in the relative *duration* of neural activity in SMA and PM, rather than activating the medial and lateral systems completely independently.

For Nachev and Husain (Nachev & Husain, 2010) "endogenous" cues are ill-defined and "intentional" actions should simply be seen as the result of a complex integration of many different external stimuli. In particular, according to Nachev and Husain, the dimension that would most clearly distinguish the two extremes is the amount of information to be integrated. Instructed actions require virtually no stimulus integration beyond recognizing the instruction, and can be accomplished in a simple, almost reflexive fashion. In contrast, decisions about the "when and what" of intentional actions may require the integration of diverse sources of both internal and environmental information, such as a complex and changing environment, internal models of the world (Berkes, Orbán, Lengyel, & Fiser, 2011), action goals (Tanji & Hoshi, 2001), memories of past actions (Hadland, Rushworth, Passingham, Jahanshahi, & Rothwell, 2001), or preceding neural activity (Soon et al., 2008). Schüür and Haggard (Schüür & Haggard, 2011) echo this criticism, and insist that considering intentional actions as the result of a complex integration of several external inputs may help escape the tendency for dualistic arguments that associate intentional actions to some sort of nonmaterial self.

In sum, the two main criticisms raised against the classical approach of comparing intentional *vs*. instructed behaviour are (1) that they differ in more than one dimension and (2) that the difference is not qualitative but rather quantitative. These criticisms have been well taken in this thesis. The working hypothesis does not consider the differences between instructed and intentional behaviour as categorical. Instead, it acknowledges that the differences exist, and seeks to empirically understand their nature and consequences. All that is required is that the intentional

and instructed cases considered are located in different positions along a continuum of degrees of freedom of selection. By dissecting the components of the differences between intentional and instructed behaviour this thesis aims at uncovering some of the obscure processes underlying "intentional" behaviour.

1. 2. 5 Neural correlates of intentional action and implications for free will

Libet *et al* (Libet et al., 1983) famously compared the time of onset of the RP with the reported time of awareness of an intention to act. In his study, participants were asked to make brisk wrist movements at their own will, and to attend to the time at which they had first experienced the will (or "urge", to use Libet's own word) to act. Participants could report the time of their awareness of the intention to act by reporting the position of a clock hand that was rotating on a screen in front of them. Libet *et al* found that the onset of the RP preceded by about 250 ms the reported time of intention, suggesting that it is not our conscious intentions that drive neural activity, but the inverse. This study has attracted a large amount of attention due to its relevance to the question of free will. Libet *et al*'s result has been widely discussed and criticized (Gomes, 2002; Trevena & Miller, 2009), but also replicated (Haggard & Eimer, 1999; Sirigu et al., 2004) and extended (Soon et al., 2008). Soon *et al* used multivariate pattern analysis classification algorithms of fMRI data to predict whether participants would freely choose to make a movement with their right or left fingers. Soon *et al* found that the volxelwise patterns of BOLD activity in the frontopolar cortex and precuneus showed above-chance predictive power up to 8 s before the time of awareness of decision.

Both Libet *et al*'s and Soon *et al*'s results suggest that unconscious brain activity builds up to give rise to conscious intentions to act, displacing the conscious self from the position of responsibility for our intentions and actions. Relating these findings to the model of "what, when, whether" decisions for action, Libet *et al*'s results provide evidence for an unconscious process in the selection of the *when* component, whereas Soon *et al*'s results reveal the unconscious determinants of *what* decisions.

Libet suggested (Libet, 1999) that between the time of awareness of intention and the time of action there is a time window where the conscious self could reclaim responsibility for our actions. Libet suggested that we may not have free will, but that we can have "free won't", and that we have the capacity of consciously "vetoing" actions that would be otherwise solely driven by our unconscious brain activity. The concept of "free won't" has consequently been a last bastion of free will. By this account, *inhibition* of action assumes a significant role in our behaviour. Of course, modern science does not accommodate such a dualistic position, and the decision to inhibit might itself have an unconscious antecedent (Hughes, Velmans, & Fockert, 2009), though these *whether* decisions have hardly been examined.

1. 3 Inhibition of voluntary action

As it has been suggested above (section 1.1), intentional behaviour emerges as a result of action and inhibition processes. Intentional behaviour, understood as behaviour that is not purely explained by responses to the environment, would be impossible without some self-control process. Often we need to hold- back from impulsive angry reactions, addictive behaviour (smoking, excessive eating, gambling), or saying or doing things that may not be appropriate in some contexts. This section will first present the two types action inhibition, namely instructed and intentional. First, the tasks used to study instructed inhibition and the main results they have yielded will be reviewed. Then, a potential distinction between instructed and intentional inhibition will be suggested on the basis of conceptual, neuropsychological and empirical evidence.

1. 3. 1 Empirical studies of action inhibition

The neuroanatomical substrates of action inhibition have largely been explored by means of two main tasks.

First, the go/nogo task has been widely used to study mechanism associated with instructed inhibition (Eimer, 1993; Pfefferbaum, Ford, Weller, & Kopell, 1985) (see

figure 1.5 A). In this task, participants are required to make quick movements to frequent go stimuli, but to withhold their movements if a nogo stimulus is presented.

Second, in the Stop Signal Task, (SST, see figure 1.5 B) quick motor actions are required in response to go signal. Crucially, on some infrequent trials stop signals appear shortly (around 250 ms) after the go signals. The precise timing of the stop signal is determined adaptively, based on the online performance of each participant. Successful stopping is therefore effortful and requires a sudden inhibitory mechanism that stops the action before it is executed.

Logan *et al* (Logan, 1994; Verbruggen & Logan, 2008a) have suggested that action and inhibition processes compete in a "race". In their model, action and inhibition are triggered by different external signals (go and stop signals, respectively), and whichever process first reaches a threshold level of neural activation will determine the action outcome. Under these assumptions, a parameter called stop signal reaction time (SSRT) can be estimated from a participant's performance (Logan, 1994). The SSRT is a precise quantifier of inhibitory function, with short SSRTs corresponding to better (quicker) inhibitory mechanisms.

The crucial difference between the SST and the go/nogo task is that in the former, adapting the SSD ensures that inhibition is often required at the latest possible moment before the action was executed. In the go/nogo task instead, inhibitory processes may be recruited earlier, and action preparation may not have been underway at the time of inhibition.

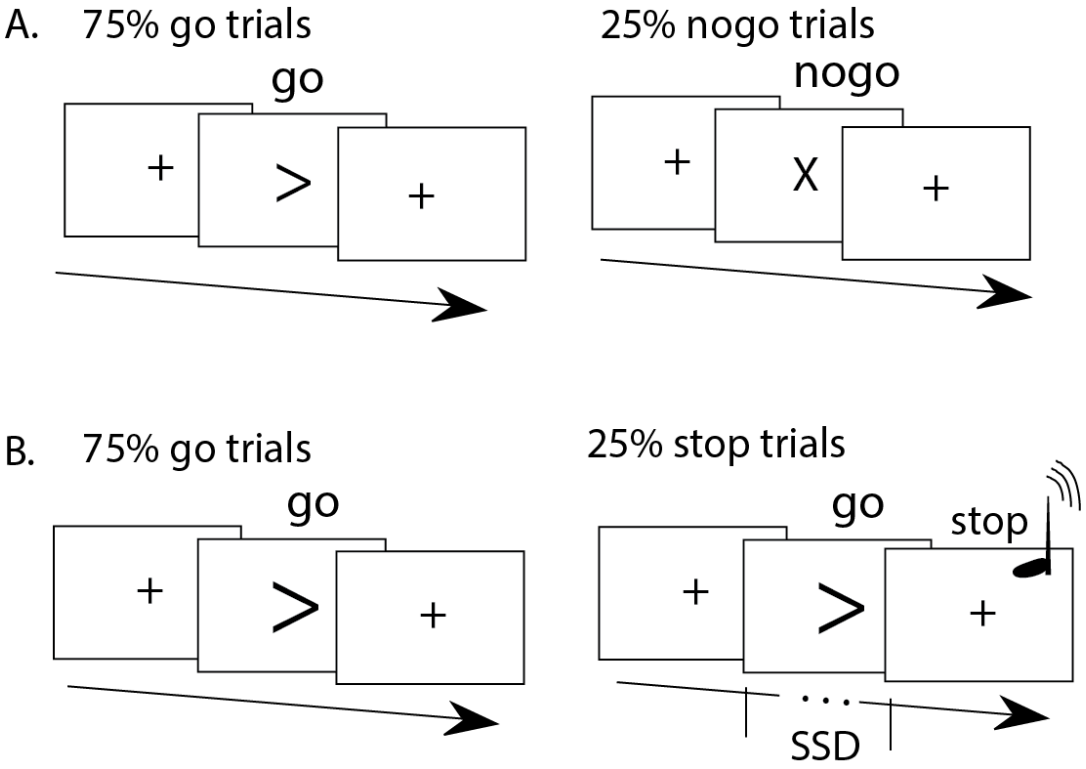


Figure 1.5 Two tasks to study instructed inhibition **A.** In the go/nogo task, a rapid succession of go signals makes acting the prepotent response. A minority of nogo (typically, but not always, 25%) signals require participants to stop their prepared actions. **B.** Stop signal task. In the majority of trials participants are required to make a quick key press in response to a go signal. In a minority of trials, a stop signal (typically a tone) will appear after the go signal. The stop signal delay (SSD) is the time between the onset of the go signal and the onset of the stop signal, and is typically varied

1. 3. 2 Neuroanatomy of action inhibition

The ability to inhibit action in the SST has been related chiefly to two cortical areas, the right inferior frontal gyrus (rIFG) and preSMA; and, subcortically, the STN within the BG. Although most studies agree that inhibitory function depends on the integrity of the three areas and their connections (Rubia *et al*., 2001), the precise division of labour between these three structures remains unclear. On the one hand, neuroimaging and direct electrical stimulation results suggest that (Aron, Behrens, Smith, Frank, & Poldrack, 2007; Swann et al., 2011) preSMA plays a role in conflict detection or resolution, and that the IFG instead is the primary structure implementing an active process of action inhibition. These results are supported by observations that increases in BOLD activity in IFG are inversely correlated with

SSRT (Aron & Poldrack, 2006), and that the extent of damage to rIFG correlates positively with SSRT. Duann *et al* (Duann, Ide, Luo, & Li, 2009), however, contests this view. The authors show (through functional connectivity analyses of fMRI data) that IFG connects to the STN only indirectly, via preSMA. On this basis, the authors suggest that preSMA is the primary inhibitory effector, whereas IFG detects a salient and behaviourally relevant signal, driving activity in the SMA which in turn implements a true inhibitory function.

In the BG, inhibitory (*stop*) processes interact with go processes. Importantly, Aron and Poldrack (Aron & Poldrack, 2006) have suggested that IFG excites STN via the hyperdirect pathway. The hyperdirect pathway sends excitatory connections from the STN to the GPi/SNr, which in turn suppresses the thalamic the output from the BG (see figure 1.2) to the cortex, and stops an action before it is executed.

1. 3. 3 Neuropsychology of action inhibition

Some neurophysiological conditions provide further evidence for the existence and importance of intentional inhibition. The anarchic hand syndrome (AHS) (Della Sala & Marchetti, 2005) is especially illustrative. AHS typically follows medial frontal and/or callosal lesions. Patients with frontal AHS will typically report that their affected hand makes compulsive, goal-oriented movements against the patient's will. These movements are clearly made in response to environmental stimuli, such as grabbing objects, food, etc. Patients are aware of the movements, and aware that they are inappropriate. Nevertheless, they are unable to intentionally inhibit them. Indeed, patients often physically restrict unwanted movements with their unaffected hand, as the only successful inhibitory strategy.

In principle, both excessive action drive and failed inhibitory mechanisms could explain AHS. Empirical evidence provides evidence for the latter being true. Cantagallo *et al* tested one AHS patient in a simple reaching task. The patient was asked to make intentional actions (i.e., pre-established sequential movements) and instructed actions (i.e., reaching towards a visual target). The authors found that the patient's affected hand was slow to initiate intentional actions *as well as* being quicker at making instructed actions (i.e., reaching towards a visual target). If the

compulsive emerged as a result of excessive action drive, then the affected hand should be quicker, rather than slower, to initiate intentional actions. Therefore, these results suggest that AHS might emerge as a result of a failure in mechanisms of intentional *inhibition*.

1. 3. 4 Principles of inhibitory control

1. 3. 4. 1 Conceptual distinction between intentional and instructed inhibition

Although the source of the differences is controversial, empirical data from diverse sources support a distinction between intentional and instructed actions, at least to the extent that they represent two extremes of a continuum (see section 1.2.2.3). This thesis explores the possibility that the same continuum may also be found in *inhibition* of action, and that an intentional/instructed distinction can be made just as clearly for action inhibition as for action execution.

Experimental studies have addressed motor inhibitory function by means of tasks that require *instructed* inhibition, most notably, the SST and the go/nogo task (see section 1.3.1). In both tasks, infrequent external signals require a rapid inhibition of prepotent simple actions. The methodological (but important) advantage of these tasks is that quick actions are the prepotent response, and inhibition cannot be planned in advance. This ensures that inhibition occurs in trials in which no action is made.

Despite the clear methodological advantages of instructed inhibition tasks, few examples of inhibitory control in naturalistic situations can be compared to a SST. Aron (2010) has suggested that motor inhibition in the SST is reactive, whereas a more ecologically valid type of inhibition is proactive. Proactive inhibition is rather related to longer term goals and motivations, and it may be targeted at inhibiting a particular response tendency. The concept of proactive inhibition can be related to intentional inhibition, i.e. the capacity to voluntarily suspend or inhibit an action, independently of a clear external signal indicating to do so. Boulingez *et al* (Boulinguez, Ballanger, Granjon, & Benraiss, 2009; Boulinguez, Jaffard, Granjon, & Benraiss, 2008) has shown that an experimental paradigm that incorporates warning

signals appearing before go signals can elicit proactive inhibition. To prevent early responding to warning signals, participants may adopt the strategy of applying a generalized suppression of reactivity that is only *released* by a warning signal. This type of inhibition is generalized, does not have a clear onset and does not follow any clear external inhibitory cues, so it may therefore be considered as intentional. Interestingly, this illustrates the ubiquity of inhibitory control. Even making simple actions may rely on inhibitory mechanisms for accurate timing.

A person may withhold an action either because of an external stop signal, or because of an intentional decision to do so. The decision to inhibit, like the decision to act, may depend on environmental stimuli (instructed behaviour), or on internal reasons and desires (intentional behaviour). For example, the current mental state may make a particular action inappropriate or undesirable, even though it might be highly appropriate in other situations. It is important to point out that the intentional/instructed dimension for inhibition is orthogonal to the intentional/instructed dimension for action. That is, one can intentionally inhibit both actions that one decided oneself to make, or actions that are tr(Stroop, 1935)iggered by environmental signals or objects. On this view, the cognitive control of action has a factorial structure. To illustrate this factorial structure, table 1.1 presents daily examples and experimental tasks for each of the possible combinations of instructed and intentional *what* and *whether* decisions.

Here it is relevant to consider contextual inhibition tasks, which typically require inhibition within a relatively complex set of rules. These tasks involve a contextual instruction to inhibit, but no overt 'inhibit' signal. For example, in Jacoby's exclusion task, participants are asked to complete a stem, e.g., "tab ", with any word apart from a word that was presented just previously. Thus, if the word table is presented first, followed by "tab " the participant must intentionally inhibit the table response, in order to achieve a correct response such as taboo (Cothran & Larsen, 2008). Similarly, the instruction to perform an antisaccade involves inhibiting a prepotent prosaccade response (Munoz & Everling, 2004), and the Stroop instruction to name the colour of a word involves inhibiting the prepotent response to read it (Stroop, 1935), and the spatial location of a stimulus may strongly influence spatially organised responses, meaning that these responses must be inhibited when spatial

parameters are irrelevant and when stimuli contain incongruent spatial information (Forstmann, Brass, Koch, & von Cramon, 2006; Simon, 1969).

Contextual inhibition therefore involves both an external stimulus, and a context which influences the way the stimulus is processed. Often, the context can be treated as a rule, for example in a set of task instructions. Thus, in Stroop tasks, the instructions specify that a word should not be read, but rather the ink colour should be named. Successful performance thus depends both on a preceding process of understanding the context, and on perceiving the stimulus. Therefore, under one possible view, applying such rules still involves sensory processing of external stimuli, but is just more complex. Therefore, contextual inhibition can be considered to be closer to instructed than to intentional inhibition.

Table 1.1: Factorial organisation of instructed and intentional control of *what* and *whether* decisions. Each cell corresponds to one of the possible combinations of instructed and intentional *what* and *whether* decisions. Each cell contains examples of daily and experimental situations that capture the processes. Adapted from Filevich et al., 2012 with permission from Elsevier Limited

		ACTION (*"what"* decision)	
		INSTRUCTED	**INTENTIONAL**
INHIBITION (*"whether"* decision)	**INSTRUCTED**	Driving towards a green traffic light, which suddenly turns to red	Suddenly cancelling a nefarious activity when realizing one is being watched
		Stop signal reaction time (Logan, Cowan, & Davis, 1984)	Pausing an action in response to external stimulation (Matsuhashi & Hallett, 2008)
	INTENTIONAL	Resisting the temptation to take another biscuit from the biscuit tin	Deciding not to send an angry email just before clicking the "send" button
		Freely choosing whether to respond to a stimulus or not (Karch et al., 2009; S. Kühn & Brass, 2009)	"Veto" task (Brass & Haggard, 2007).

In the first place, action and inhibition may occur in an automatic way, as in the case of responses to sudden simple environmental events such as traffic lights. These kinds of situations are well operationalized by classical instructed inhibition tasks, such as the go/nogo task, where participants are required to act in response to go stimuli but not to nogo stimuli. But inhibition may also occur without any external instructions, in order to control or limit a natural tendency to respond to impulses.

An example would be to resist from succumbing to potential addiction, where actions may be initiated automatically but a second thought will control the impulse. Empirical tasks operationalizing this intentional inhibition of instructed action may include free decisions about whether to follow a default action plan or not.

On the other hand, inhibition of intentional action may occur as a quick reaction to an unexpected event in the external environment. In daily life, such situations occur when an admittedly "wrong" action is suddenly interrupted by an external noise. A task that captures this process may instruct participants to act freely but stop their actions if they hear an external tone. And finally, intentional inhibition of intentional action may take place in everyday life when we "change our minds" about making actions that we had freely selected. Tasks that operationalize this process will typically allow participants to make *when* or *what* decisions, and additionally ask them to make a final whether checkpoint decision.

Intentional inhibition shares some features with instructed inhibition. For example, there is in both cases a prepotent or otherwise salient motivation for action. Further, the preparatory processes that lead to action are already underway when inhibition occurs. However, intentional inhibition has features that are not shared with other forms of inhibition. By definition, the process or signal that cancels or inhibits the action is not the result of any external signal or instruction, but is crucially generated *internally* by the participant herself. In this respect, intentional inhibition clearly differs from classic psychological paradigms where an external stop signal is used to trigger inhibition (Logan & Cowan, 1984), or NoGo tasks (Pfefferbaum *et al*., 1985; Eimer, 1993).

1. 3. 4. 2 Methodological difficulties in isolating intentional inhibition

A methodological difficulty in the study of intentional inhibition arises because an experimenter needs to assume that *something* was inhibited. That is, she will hope that there was a process which would have lead to action had it not been inhibited. But, in the absence of any behaviour, what evidence is there that an action would

have occurred? Given the absence of overt behaviour, neuroimaging methods provide a particularly useful approach. The following section describes two neuroimaging studies that support intentional inhibition as a distinct process. In particular, three situations may involve very different neural processes but are behaviourally identical (in the sense that neither produces any behavioural output) and will therefore be difficult to distinguish, namely instructed inhibition, intentional inhibition and intentional early decisions not to act.

First, although theoretical grounds may suggest a possible distinction between intentional and instructed inhibition, empirical data are necessary to provide stronger support. Second, as was discussed above (see section 1.2.2) two possible decisions may lead to an absence of overt behaviour, namely early *whether* decisions not to act and later *whether* decisions to inhibit action. To study intentional inhibitory processes, these two situations should be clearly distinguished

1. 3. 4. 3 Neuroimaging evidence for a distinction between intentional and instructed inhibition

Only a limited number of experimental studies have addressed a process of intentional inhibition. Brass and Haggard (Brass & Haggard, 2007) asked participants to intentionally prepare and execute a simple keypress action on some trials, but on other trials to prepare the action and then withhold it at the last possible moment. Participants freely chose on each trial whether to act or inhibit. Participants reported the time of their intention to act, even on trials where no action in fact occurred, and this was used for event-related fMRI comparisons between action and inhibition conditions. The contrast of inhibition vs. action trials revealed BOLD activity in the dorsal fronto median cortex (dFMC). In addition, the analyses revealed a significant correlation between each participant's percentage of inhibited trials and inhibition-related activity in dFMC.

Kühn *et al* (Kühn, Haggard & Brass, 2009) asked participants to freely decide between executing and inhibiting a keypress action. Their task provided a prepotent external drive to act, as the action of pressing the key would avoid an unpleasant sound. Some trials consisted of external instructions to either perform or inhibit the

keypress. Other trials allowed participants to freely decide what they would do. Intentional inhibition was identified by contrasting trials involving a voluntary decision to inhibit with trials involving a voluntary decision to proceed with the prepared action. This contrast revealed BOLD activity in dFMC, close to the area reported by Brass and Haggard (Brass & Haggard, 2007), though slightly more ventral. As in Brass and Haggard's earlier study, the authors found a correlation between individuals' probability of inhibition and inhibition-specific BOLD activity. Kühn *et al*. (Kühn, Haggard & Brass, 2009) argued that, given that dFMC appears in the contrast between deciding to inhibit and deciding to act, it cannot be related exclusively to the decision itself. Instead, the authors show through connectivity analyses that the RCZ is responsible for the decision, while the dFMC simply expresses the decision outcome "inhibit".

The clear anatomical substrate for intentional inhibition helps to differentiate it from other forms of inhibition, and inhibition-related processing. First, instructed inhibition in SST has been associated with two quite different areas, the right inferior frontal gyrus, and the SMA. A recent large meta-analysis of the SST (Swick, Ashley, & Turken, 2011), included the contrasts Stop > Go, Stop > baseline and Successful stop > Unsuccessful stop. The resulting activation likelihood estimation (ALE) map contained major clusters in the left insula, extending into thalamus and putamen; the posterior cingulate (BA 23); right insula, extending to inferior precentral gyri (BA 9) and the superior frontal gyrus (medial BA 6, including the pre-SMA), the right middle frontal gyrus (BA 9), and the right inferior parietal lobule (BA 40). This analysis identifies a network involved in instructed inhibition, focused on a lateral and a medial frontal cluster. Importantly, the identified network for external inhibition does not overlap with the medial prefrontal areas associated with intentional inhibition. In particular, the medial cluster for internal inhibition is clearly posterior to the medial cluster for intentional inhibition (Filevich, Kühn, & Haggard, 2012a) (see figure 1.6).

1. 3. 4. 4 Neuroimaging evidence for a distinction between intentional inhibition and intentional nonaction

To confirm the difference between the BOLD correlates of action selection and those of intentional inhibition, a recent study (Filevich et al., 2012a) has shown through ALE analyses that the brain areas typically associated with action selection (derived from a contrast of intentional action > instructed action or intentional action > rest) do not match those associated with intentional inhibition. Intentional action selection is typically associated with increased BOLD activity in preSMA and SMA, areas different from the more anterior activation of dFMC shown to be associated with intentional inhibition (Brass & Haggard, 2007) (see figure 1.6). This analysis suggests that the activation associated with early selection of voluntary actions is distinct from the activation associated with late intentional inhibition of actions that are already prepotent.

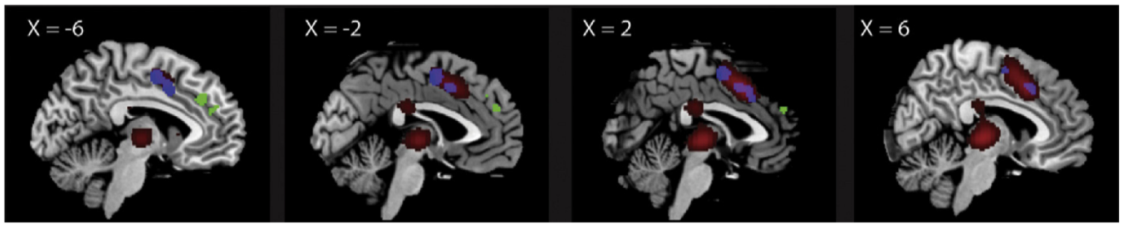


Figure 1.6 Results of ALE meta-analysis from 21 stop-signal studies (red) (Swick et al., 2011), 11 response selection studies (blue) and 7 intentional inhibition studies (green). Note the clear distinction between the more posterior preSMA coordinates of stop-signal and response selection studies and the more anterior dmPFC coordinates of intentional inhibition studies. Taken from Filevich *et al*., 2012 with permission from Elsevier Limited

1. 3. 4. 5 Speculative neurostimulation evidence for a distinction between intentional and instructed inhibition

A rather different source of evidence for a distinction between intentional and instructed inhibition comes from the neurosurgical literature. In cases of presurgical evaluation in drug-resistant epilepsy, subdural electrode arrays may be placed on the cortical surface. Each electrode of the array can be stimulated individually to assess the function of the underlying local area of cortex. Similar stimulation techniques

can be used intraoperatively. This method has been exploited by Penfield and Welch (Penfield & Welch, 1951) to find for the 'positive' sensorimotor signs evoked in specific muscles, leading to the famous motor homunculus.

Interestingly, the neurosurgical literature also identifies electrode sites where direct electrical stimulation causes *slowing* or *suppression* of ongoing movements (Lüders, Dinner, Morris, Wyllie, & Comair, 1995). If a patient is asked to perform rapid alternating eye, tongue, hand or foot movements, and an electrode is stimulated while the movement is ongoing, a negative motor response may be found. These sites have been termed 'negative motor areas' (NMAs) (Lüders *et al*. 1995). Alternative simple explanations of non specific motor arrest, such as loss of consciousness, can be excluded. Since the 1950s, over 20 studies have reported NMAs upon direct cortical stimulation. The total frequency of NMAs varies dramatically between studies, perhaps reflecting the difficulty of extensive and comprehensive sampling given the strict clinical restrictions of this unique setting.

Clearly, such external stimulation will bypass any decision to inhibit on the patient's part, so it can say little about the natural circumstances under which this suppression occurs. On the other hand, stimulation offers a well-controlled method that can reveal how the suppressive mechanism functions.

Previous discussions of NMAs have been largely confined to the neurosurgical literature. The general interpretation in that literature suggests that the normal function of NMAs is the fine regulation of motor output (Chauvel, Rey, Buser, & Bancaud, 1996; Ikeda et al., 2009; Mikuni et al., 2006). An alternative interpretation is that NMAs reflect a functional system for inhibition of action. Given the widespread neuropsychological consensus that inhibition of action is a crucial aspect of both cognitive control of behaviour, this interpretation would make NMA data highly relevant to cognitive neuropsychology. A detailed review of the NMA literature with a specific emphasis on the possible contribution of NMAs to inhibitory processing is beyond the scope of this introduction, but in the following section a possible functional inhibitory role for NMAs is discussed.

Several lines of evidence suggest that NMAs may be functionally relevant for inhibition (Filevich, Kühn, & Haggard, 2012b). First, stimulations on many sites that

produce positive motor effects do not also produce negative motor responses. In fact, highly complex sequences of functional action can be evoked by some instances of electrical stimulation (Bancaud *et al*., 1976), yet these positive motor effects can be readily dissociated from negative motor effects. Second, NMAs are sometimes found in quite different areas from positive motor areas, and usually anterior to positive motor areas (Uematsu *et al*., 1992). Third, NMA localisation matches the areas showing increased BOLD activity associated with response inhibition in stop signal tasks (see review articles by (Chikazoe, 2010; Levy & Wagner, 2011; Swick *et al*., 2011). Fourth, NMAs are sometimes found at lower intensity than positive motor effects (Mikuni *et al*., 2006). Taken together these findings suggest that negative motor responses do not simply arise from disrupting normal physiological activity in excitatory areas.

Finally, in one recent study the roles of preSMA and IFG (Swann *et al*., 2011) were addressed in a rare patient with implanted electrodes over both preSMA and IFG. In this case study of a single patient, evidence from connectivity patterns, functional properties and direct electrical stimulation suggest that NMAs may play a functional role in motor inhibition, and that they may do so by driving a network of several frontal cortical areas that provide a balance between excitation and inhibition.

Interestingly, NMAs have been reported in two distinct clusters: a medial cluster focussed on the SMA, and a lateral cluster focussed on the IFG and premotor cortex (see figure 1.7). In relation to a possible distinction between intentional and instructed inhibition, a speculative account suggests that the medial NMA cluster might be involved in processes of intentional inhibition, whilst lateral NMAs may be involved in instructed inhibition processes. The available empirical data is based on clinical mapping results, where control is applied externally by the experimenter, and cannot be said to be truly intentional – even if it involves activation of areas whose normal functions include intentional action or intentional inhibition. Therefore more research using tasks that require intentional inhibition may be necessary to test this hypothesis.

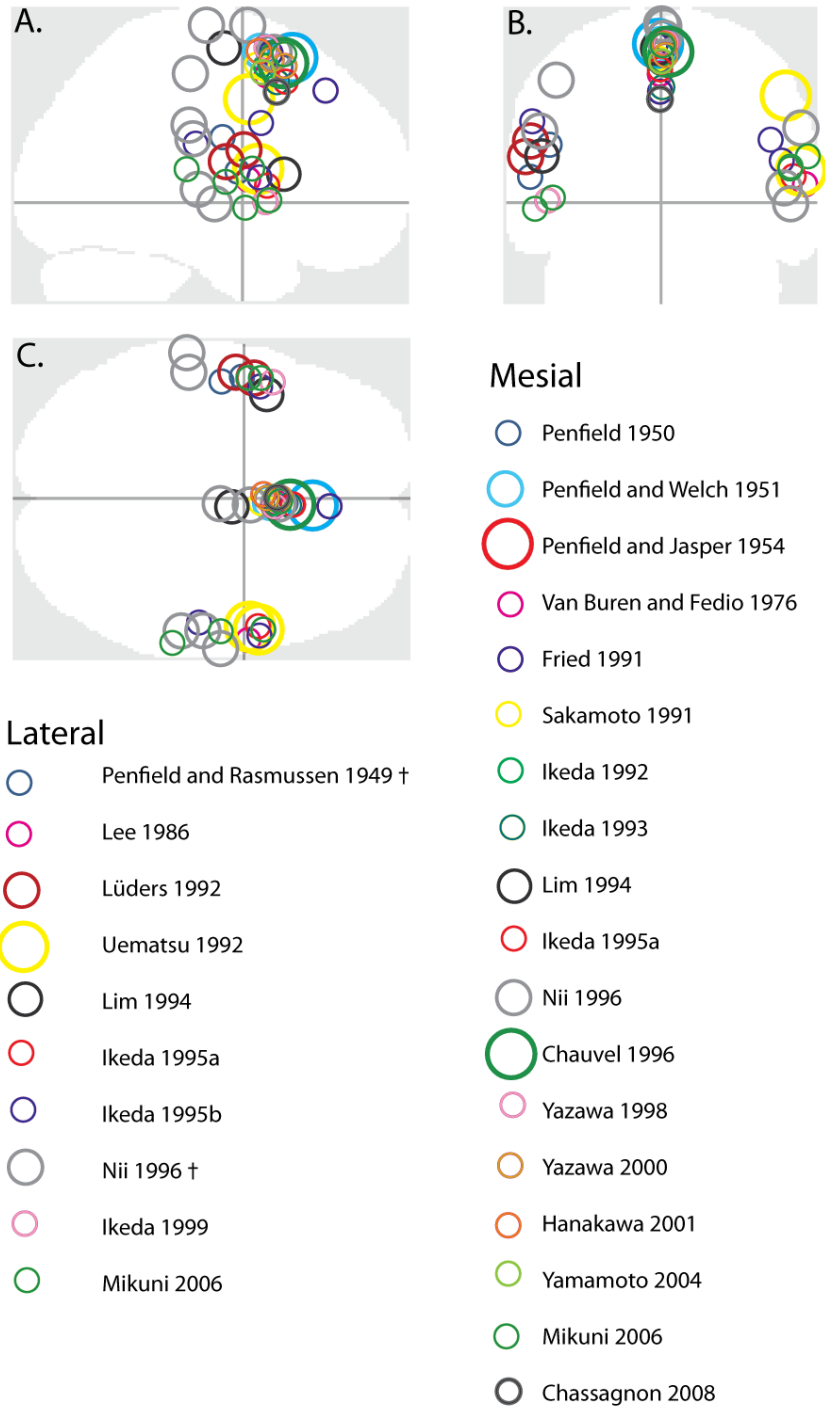


Figure 1.7 Approximate location of reported NMAs shown on a transparently rendered glass brain. **A** Sagittal view, **B** coronal view, **C** axial view. Coordinates were approximated by visual inspection of the original figures. Small circles represent 1-5 NMA sites, medium circles represent 6-20 NMA sites, and larger circles represent >20 NMA sites. Different colours represent individual studies, but colours may be repeated due to one individual study showing more than one NMA cluster. † Indicates studies in which the lateralization on the NMAs was not reported, and were therefore inferred (Nii et al.,1996) or depicted on the right (Penfield and Rasmussen, 1949). Gray lines intersect at the anterior comissure. Adapted from Filevich *et al*, 2012b. Reproduced with permission of Elsevier Limited.

1. 3. 5 Electrophysiology of action inhibition

One of the advantages of externally triggered inhibition is that the stopping process, when it occurs, is precisely localized in time and the electrophysiological signature of inhibition may be studied. EEG studies in SST and go/nogo tasks have therefore focussed on instructed inhibition and provided evidence for the involvement of a complex ERP component that is identified as N2/P3 (Falkenstein, Hoormann, & Hohnsbein, 1999; Ramautar, Kok, & Ridderinkhof, 2004). The N2/P3 complex presents greater amplitudes for nogo trials as compared to go trials (Jodo & Kayama, 1992) and for successful nogo as compared to unsuccessful nogo trials (Kok, Ramautar, De Ruiter, Band, & Ridderinkhof, 2004).

The N2 component peaks at around 200 ms after nogo signal onset and is maximal at the frontal electrodes. Donkers *et al* (Donkers & van Boxtel, 2004) designed a "go/GO" task, in which the infrequent GO stimuli did not require action inhibition but a stronger action instead. The authors found an N2 component associated with GO stimuli, as well as when participants did the classical go/nogo task. Therefore the N2 may be related to the detection of conflict or salient and behaviourally stimuli rather than with inhibition *per se*.

The P3 component peaks at around 300 ms and has a fronto-central distribution. It appears to be directly involved in stopping processes and peaks earlier for successful stop trials as compared with unsuccessfully stopped ones (Bekker, Kenemans, Hoeksma, Talsma, & Verbaten, 2005; Dimoska, Johnstone, & Barry, 2006). To address the relevance of P3 for motor inhibition, Smith *et al* (J. L. Smith, Johnstone, & Barry, 2008) compared a classical go/nogo task with an equivalent "count/no-count" task, in which participants were asked to covertly count the number of frequent go trials, but not to include the infrequent nogo trials. In this way, saliency and behavioural relevance effects were controlled for. The authors found that the differences in the amplitudes of the P3 component were larger between go and nogo trials that between count and no-count trials, suggesting that this component is indeed related to action inhibition.

It has been suggested that action inhibition critically relies on conscious processes and can therefore only be triggered by supraliminal stimuli (Dehaene et al., 2003; Mayr, 2004). However, Hughes *et al* (Hughes et al., 2009) showed that unconscious go/nogo primes can influence both behavioural performance in a go/nogo task and the amplitude of the N2/P3 complex. The authors showed that nogo primes significantly reduced the amplitude of the N2/P3 complex, suggesting that some priming of inhibitory function occurred following the presentation of subliminal stimuli, and thus reducing the inhibitory effort necessary to stop an action. In a similar vein, Van Gaal et al (van Gaal, Ridderinkhof, van den Wildenberg, & Lamme, 2009) have shown that subliminal primes associated with stop signals in the context of a SST can facilitate action inhibition.

Fewer studies have addressed the electrophysiological correlates of intentional inhibition. In one recent study, Walsh *et al* (Walsh, Kühn, Brass, Wenke, & Haggard, 2010) tested participants in a Libet task. In addition, participants were asked decide to inhibit their actions in some trials at the very last moment. If they had inhibited their action, participants were asked to report the time of inhibition. Walsh *et al* time-locked the EEG recordings to the reported time of inhibition, and found a frontally-distributed event-related synchronization in the beta band peaking at 12 ms before the reported time of inhibition. The authors suggest that this may be an electrophysiological signature of intentional inhibition.

1. 4 Outstanding questions in volitional control of behaviour

A vast body of literature points to a distinction between intentional and instructed action. Convergent evidence from phenomenological accounts, behavioural data and neural data supports the validity of an instructed/intentional distinction in action processes.

However, the nature of this distinction remains unclear and is still hotly debated. The scientific value of a theory depends on its ability to provide simple explanations of phenomena other than those that it was originally devised to explain. This thesis therefore attempts to extend the intentional/instructed distinction in three important directions that remain unexplored.

First, if the intentional/instructed distinction is valid for the control of behaviour as a whole, it should not be restricted to action only. Given the importance of inhibition of action for the control of behaviour, it may be possible to find empirical support for a distinction between intentional and instructed *inhibition*. Because of the methodological difficulties associated with studying inhibition of action, this issue has remained relatively unexplored up to now.

Second, if the instructed/intentional distinction is a relevant one for action *control*, it should presumably also have an impact on the experience of action, given that people are generally aware of their actions. The operational distinction between instructed and intentional action has been scientifically useful because it helped to bypass subjectivity, and is based on information processing. Perhaps because of the methodological difficulties typically associated with incorporating subjective experience in empirical research, this aspect of intentional behaviour has been largely neglected. Arguably however, it lies at the core of intentional action. This thesis therefore aims at exploring the instructed/intentional distinction with an emphasis on the subjective experience associated with it.

There are clear methodological problems associated with the study of both intentional inhibition and the subjective experience associated with intentional behaviour. Consequently, this thesis describes several different methodological approaches that are suitable to overcome the practical and conceptual difficulties in the study of intentional behaviour.

This thesis studied intentional behaviour and what distinguishes it from instructed behaviour. It examined the neural processes that range temporally from decision

making to the processing of decision consequences. The thesis is organized to highlight three crucial aspects of the instructed/intentional distinction that had not previously been explored. First, three experiments explored a possible distinction between instructed and intentional inhibition of action, and its functional relevance. Second, two experiments examined the period between response selection and response execution, i.e. the period of decision maintenance. And finally, two experiments focussed on the elusive phenomenology of intentional action.

In particular, and as a first inferential step, chapter 2 explored potential differences in the subjective experience associated with intentional and instructed inhibition of action. Preparatory processes for action and action inhibition were studied through their effects on time perception. Results revealed that whereas instructed action and instructed inhibition were associated with clearly different degrees of time compression, this was not the case for intentional action and intentional inhibition. In the case of intentional behaviour, action and inhibition trials were indistinguishable on the basis of their effects on time compression. Intentional conditions led to "intermediate" levels of time compression as compared to their instructed counterparts.

Then, to explore the issue of intentional inhibition in more naturalistic settings, the experiment described in chapter 3 aimed to induce in participants strong urges to act, and therefore strong requirements for action inhibition. In this case the neural processing of the sensory consequences of behaviour was compared for instructed and intentional inhibition. Results showed that the neural processing of the consequences of intentional decisions to inhibit action differ from those of instructed decisions to inhibit action. The differences between intentional decisions to act and inhibit are smaller that between instructed decision to act and inhibit. Therefore, in line with the results from chapter 2, the processing of the consequences of intentional decisions represented "intermediate" situations between the extreme instructed decisions, perhaps because intentional decisions always admit the "possibility of doing otherwise".

Because of the special status of intentional inhibition regarding the conscious control of impulsive action, chapter 4 investigated the relation between the neural processes underlying intentional inhibition and conscious awareness. Specifically, the

experiment investigated whether decisions to inhibit could be influenced by unconscious preceding neural activity. Results revealed that, in the absence of strong biases for any of the possible response outcomes, intentional decision processes may capitalize on the intrinsic fluctuations of neural activity to bias the outcome. This may be a potential mechanism whereby intentional decisions to inhibit action can select between two equivalent response alternatives.

If voluntary behaviour involves neural structures associated with intentions, one intuitive account would hypothesize that intentional decisions are stronger, and more difficult to change, than instructed decisions. However, results from experiments in chapters 2 and 3 suggested the opposite scenario. Therefore, in a second extension of the instructed/intentional distinction, two experiments explored potential differences in the period between action preparation and action execution, by addressing the *strength* of intentional decisions. The study presented in Chapter 5 explored the flexibility of neural representations of intentional response alternatives. The results showed that the internal representation of actions is not flexible, and presents a lag relative to sudden changes in the external environment.

Chapter 6 explored the penetrability of intentional decisions by external task-irrelevant distracters. The results did not show differences between instructed and intentional action in the susceptibility to external distraction, in keeping with other findings previously reported in this thesis.

Two final experiments directly explored the subjective experience associated with intentional action. Chapter 7 explored the timing of the emergence of conscious intentions to act, in cases of instructed or intentional action. Results did not show differences in the timing of conscious awareness of intention for different action types. These results stressed the methodological difficulties in addressing the subjective experience of intentional action.

Chapter 8 addressed the subjective feeling of acting intentionally, and attempted to identify the neural structures that may underlie it. The results suggest that a postcentral area that had not been related to intentional action before may be responsible for the emergence of the feeling of acting freely.

Together, this thesis provides strong and consistent evidence for a validity of a distinction between instructed and intentional action, the distinction is not restricted to the single aspect of action generation for which it was originally developed. Rather, the work presented here suggests that the distinction can be usefully be extended both to the subjective experience of action execution, and also to the control of action inhibition.

Chapter 2 Intentional action and intentional inhibition: effects on time perception

Making and inhibiting actions are associated with characteristic subjective experiences. Are these experiences comparable for instructed and intentional behaviour? Biases in subjective experience of time following intentional or instructed action or inhibition were measured in a factorial design in a psychophysical experiment. Biases in time perception were smaller for intentional actions than for instructed actions, but were greater for intentional inhibition than for instructed inhibition. Moreover, intentional action and intentional inhibition produced similar compressions of subjective time. These results suggest that intentional inhibition is a distinctive cognitive operation with widespread experiential effects.

2. 1 Introduction

Intentional and instructed decisions produce virtually identical behavioural outcomes from the point of view of an external observer. In cases of action inhibition, instructed and intentional processes are associated with no overt behaviour at all. Consequently, behavioural measures have limited value, and other approaches are necessary to reveal differences between intentional and instructed decisions. Typically, neurophysiological measures have been used (see the introduction of this thesis).

As a possible alternative approach this study considered the subjective experience associated with intentional and instructed decisions. If intentional decisions involve different mechanisms from those of instructed decisions, each process will in turn lead to distinct subjective experiences. Therefore, as a first inferential step, to explore differences between intentional and instructed decisions in cases of both action and inhibition, the subjective experience associated with these processes was addressed.

Subjective time duration is a candidate aspect of subjective experience that has been linked to action processes and is readily measurable. Accurate and precise time estimation is important to synchronize action with dynamic events in the external environment, and to interpret causal relationships based on temporal consistency.

Experimental research has addressed the relationship between action and time perception (see figure 2.1). In a typical action and time perception task, participants are asked to make actions and concurrently to estimate the duration of a "test" interval that occurs around the time of their action. Perceived durations of the test interval may be compared with other "probe" intervals. Using these paradigms, time dilation effects or *chronostasis* have been consistently found after the time of action. A vast body of literature shows that making actions leads to biases in time perception in the context of these paradigms. Different actions, such as saccades (Yarrow, Haggard, Heal, Brown, & Rothwell, 2001), key presses or voice commands (Park, Schlag-Rey, & Schlag, 2003) and arm movements (Yarrow & Rothwell, 2003) can lead to chronostasis. This effect is generally taken to be a compensatory

mechanism that can correct for the perceptual "time lost" during action, or for more general mechanisms such as movement preparation and attention (Yarrow, Whiteley, Rothwell, & Haggard, 2006).

However, it has recently been shown that time dilation associated with action preparation does not only emerge as a compensation for sensory suppression (Hagura, Kanai, Orgs, & Haggard, 2012). In a series of elegant experiments, Hagura *et al* have shown that action preparation is associated with dilation of subjective time, whilst at the same time leading to enhanced sensory perception. The authors asked participants to make reaching movements to targets presented on a screen, and to judge the duration of a target interval presented immediately before their movement. Interestingly, the authors found that when participants knew the exact spatial location of the target, and hence could precisely prepare their actions, the time dilation effects were maximized. Time dilation effects were smaller when participants were uncertain about the precise location of the reaching target. These results suggest that biases in time perception may be tightly linked to the advancement of action preparation present at the time of occurrence of the interval to be judged.

The opposite effect of time dilation, namely time *compression*, has been found in periods around the time of action. Morrone *et al* (Morrone, Ross, & Burr, 2005) evaluated the effects of saccadic eye movements on time perception. Three participants were asked to estimate the duration of a time interval occurring between the appearances of two pairs of horizontal bars presented in the periphery of the visual field. Two time intervals were presented in each trial, namely the test interval and the probe interval. The test interval had a fixed duration and was presented at unpredictable times relative to a saccade onset. The probe interval had a variable duration, and was presented well after the saccade. Participants were asked to judge whether the test interval had been longer or shorter than the probe interval. The results showed that time intervals around the time of saccades were consistently perceived as shorter than they had actually been. This chronostasis effect did not depend on the size of the saccades or on the peripheral distance of the visual stimuli (the horizontal bars) to the saccadic target. The time compression effect was greatest over a period of ± 200 ms relative to saccadic onset.

Other research has shown that time compression effects are not restricted to saccadic movements. The "intentional binding" effect described by Haggard *et al* (Haggard et al., 2002) shows time compression after finger movements (both instructed and intentional; see section 1.2.2.2). This effect appears to be related to the conscious motor intentions occurring prior to action, and not to processes of intentional action selection.

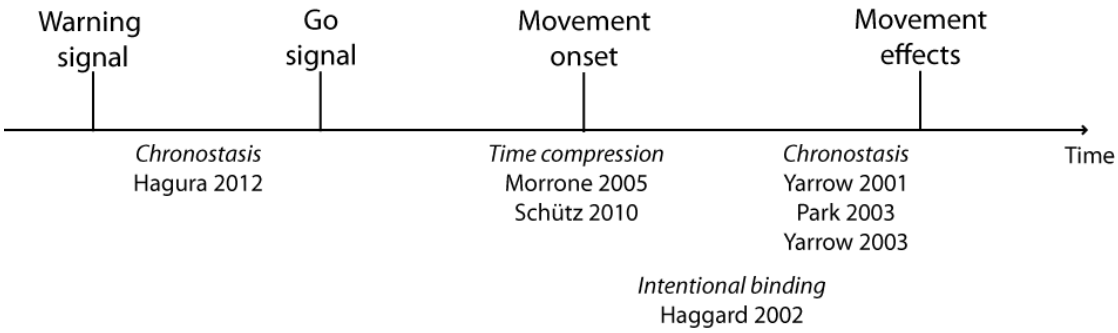


Figure 2.1 Some of the temporal illusions described in relation to action

In addition to action-related processes, other cognitive processes may influence time perception. In the well documented "filled duration illusion" (E. Thomas & Brown, 1974), an interval during which a series of discrete perceptual events occur is typically judged to be longer than another interval of equal length in which no perceptual events occur. For example, if during only one of two intervals of equal length a series of auditory tones are played, that "filled" interval will be perceived as having a longer duration than the "empty" interval during which no other tones were presented.

In a recent large meta-analysis of experimental studies on time perception, Block *et al* compared time estimation biases between conditions of high and low cognitive load (R. A. Block, Hancock, & Zakay, 2010). The authors distinguished between prospective and retrospective paradigms. In prospective paradigms, participants know beforehand that they will be required to judge the duration of an interval. In

retrospective paradigms, participants are asked about the duration only after they have been exposed to the time interval. The meta-analytic results of Block *et al* reveal that in prospective time estimation paradigms, higher cognitive load conditions are associated with stronger time compression effects. This suggests that cognitive mechanisms such as attention and working memory may impair time perception.

Given the effects of both action and cognitive processing on time perception, the two experiments reported here explored potential effects of action inhibition and intentional decisions on time perception. In particular, it was hypothesized that inhibited actions can lead to biases in time perception; and that the additional decision making processes required in intentional decisions would lead to differential biases in subjective time perception. If the additional cognitive process of intentional decisions leads to a "filled duration" effect, then intentional decisions will presumably be associated with stronger time dilation effects than instructed processes.

To test the hypothesis, a dual paradigm was designed. The paradigm included on the one hand, a go/nogo task that could be either instructed or intentional. That is, the outcome of the go/nogo task could either be specified in the instructed conditions, or depended on the participants' choice in the intentional conditions. On the other hand, the go/nogo task was combined with a temporal judgement task. In a Michotte-type paradigm (Michotte, Thines, & Crabbe, 1991; Yantis, 1995), participants were asked to estimate the duration of a visual event that was independent of the participant's response. In this way, the effects of the intentional go/nogo task on the time perception task could be determined.

Experiment 1 revealed differences in time perception between instructed and intentional decisions. Experiment 2 showed that these differences are in turn associated cognitive processing that occurs prior to the execution of a response.

2. 2 Experiment 1

2. 2. 1 Methods

Twenty naïve participants (4 female, mean age 22 years) took part in the study. All participants had normal or corrected to normal vision. Procedures were approved by the UCL research ethics committee and were in accordance with the principles of the Declaration of Helsinki.

Experimental stimuli were displayed on a CRT monitor with a refresh rate of 60Hz. Participants sat 60 cm away from the screen, and rested their right index finger on an "action key" connected via serial port to the stimulus computer. Their left index hand rested on the arrow keys of a standard keyboard.

The experimental paradigm consisted of two simultaneous tasks, indicated by two visual cues displayed on the screen (see below). First, in a go/nogo task, participants were asked to make or inhibit quick key presses. Second, in a temporal judgement task, participants were asked to make indirect temporal judgements. Covert cognitive processes presumably occurring during the go/nogo task were addressed through their effects over performance in the temporal judgement task. Participants were asked to favour both speed and accuracy in the go/nogo task, and to favour accuracy over speed in the temporal judgement task.

Go/nogo task

Stimuli were presented over a grey background. Each trial developed as follows. Participants initiated each trial by pressing the action button. A small white cue (0.1 degree of visual angle) was shown revolving around a white central fixation point (see figure 2.2), with a radial distance of 1.5 degrees and at a fixed angular speed of 0.5 revolutions per second. The direction of rotation (clockwise or counter clockwise) was fully randomized across trials. Participants were asked to maintain fixation on the central cue, and to avoid pursuit of the peripheral cue. Further, they were asked to prepare a right key press at the onset of the trial. After a variable period of between one and two full revolutions, the peripheral cue disappeared for a

fixed period of 700 ms (occlusion period). Simultaneously with the occlusion period onset, the central cue changed colour. The central cue could turn green, red or yellow. The colour of the central cue, together with the participant's behaviour, determined the trial type. Each trial fell in one of five possible conditions: instructed go, instructed nogo, intentional go, intentional nogo and baseline condition. A green central cue instructed participants to make a quick key press (instructed go condition). A red central cue instructed participants to inhibit their prepared action (instructed nogo condition). A yellow central cue allowed participants to decide whether to act (intentional go) or inhibit their key press (intentional nogo). Participants were asked to roughly balance the number of intentional go and nogo trials, but to try to avoid responding in a prespecified fashion. The four experimental conditions fell into a factorial design, with the factors of source (instructed/intentional) and outcome (go/nogo).

Further, in a baseline condition, the central cue was coloured blue from the onset of the trial. Unlike in the four experimental conditions, in the baseline condition participants were asked not to prepare the key press at all, and could concentrate on the temporal judgement task. The baseline condition therefore measured participants' subjective time perception in the absence of any action-related processes.

Temporal judgement task

The temporal judgement task was adapted from previous studies requiring time estimation (Parkinson, Springer, & Prinz, 2011). After the 700 ms occlusion period, the peripheral cue reappeared and continued its normal revolution for a period of 800 ms. Crucially, angular offsets were introduced to the angular position of reappearance of the peripheral cue, introducing a mismatch between the expected and actual positions of reappearance. The maximum angular offset introduced corresponded to a temporal offset of ±700 ms of duration of the occlusion period. Thus, on reappearance, the peripheral cue could be seen either too far ahead or too far behind the "correct" position of reappearance, estimated on the basis of the original travelling speed and duration of the occlusion period. Participants were then asked to make a two-alternative forced-choice judgement to indicate whether the

peripheral cue had travelled too far or not far enough during the occlusion period. The temporal offset was adaptively adjusted with a QUEST algorithm (Watson & Pelli, 1983). The data were then fitted with a cumulative binomial function, to approximate the psychometric function. The point of subjective equality (PSE) was then measured as the 50% point of the cumulative function fitted to the psychometric function. Normally distributed noise around the PSE were added to the temporal offset suggestions produced by the QUEST algorithm to improve the estimation of the psychometric functions. To further improve the estimation of the psychometric curve, only extreme values of angular offset were presented in the first 80 trials.

Just-noticeable differences (JNDs) were calculated as half of the difference between the 75% and 25% thresholds of the cumulative function fitted to the psychometric data for each condition and each participant.

Leftwards shifts (towards smaller values) of the psychometric curves would indicate time *compression*, as it would imply that participants overall perceived the peripheral cue to have travelled *shorter* distances. Because the peripheral cue speed was constant, shorter expected distances would imply occlusion periods perceived as shorter. In turn, rightwards shifts (towards larger values) of the psychometric curves would be indicative of time dilation, because longer expected travelling distances implied that elapsed periods were perceived as longer.

A simple cognitive framework was used to characterize the cognitive processes that took place in each experimental condition (see figure 2.5). In all conditions except from baseline, participants would have made a perceptual decision about the colour central cue within the duration of the occlusion period (see figure 2.2). Additionally, in intentional trials participants were required to make an intentional decision whether to execute the action they had prepared. The occlusion period could not be synchronized exactly with any single cognitive process. The 700 ms occlusion period was expected to encompass both perceptual and intentional processes, plus most actions or inhibitions. The exact time of actions (go RT) could be measured, but the "nogo RT" could only be inferred from the go RTs. It was assumed that the distributions of go and nogo RTs would be similar for each participant. Finally, some of the go and nogo RTs were expected to fall within the occlusion period.

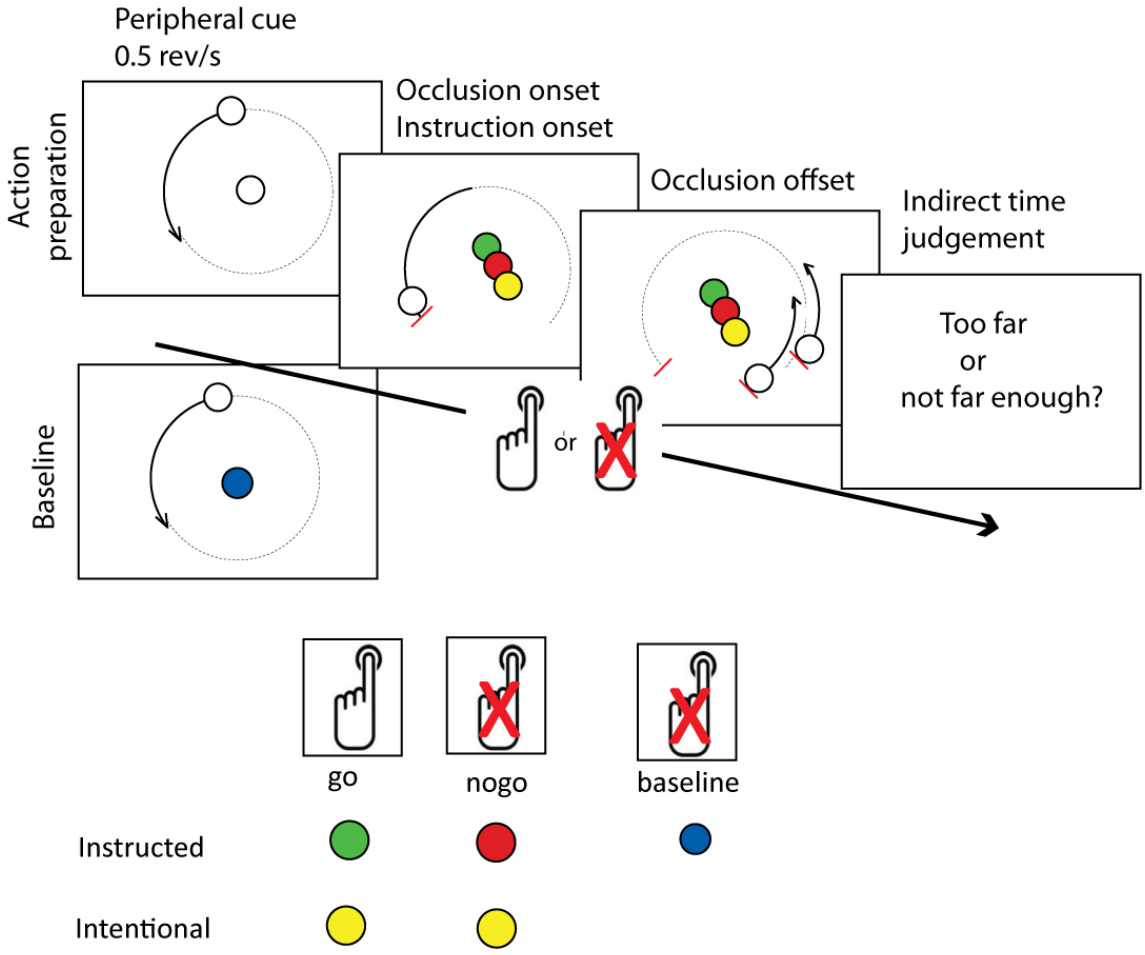


Figure 2.2. Experimental task for experiment 1. A peripheral cue revolved around a central cue at a fixed speed (0.5 revolutions/s). Both cues were white dots displayed over a grey background. Each trial included two concurrent tasks, a temporal estimation task (given by the peripheral cue) and a go/nogo task (given by the central cue). For the temporal estimation task, the peripheral cue rotated for a variable period of between one and two full revolutions, until it disappeared for a fixed time of 700 ms (occlusion period). The peripheral cue then reappeared and continued revolving for a further 800 ms. The position of reappearance of the peripheral cue could either be too far ahead or lagging behind the correct reappearance position, given the fixed speed of the peripheral cue and length of the occlusion period. Short red lines on the circle indicate the onset and offset of the occlusion period. Participants were asked to make an indirect temporal judgement by indicating whether they perceived that the cue had travelled too far or not far enough. For the go/nogo task, at the time of disappearance of the peripheral cue, the central cue changed colour, to indicate whether participants should make a key press, inhibit their prepared action or intentionally decide in the go/nogo task. In the baseline condition, a central cue coloured blue from the onset of the trial indicated to participants that they should not prepare their action at all, and should concentrate exclusively on the temporal judgement task.

The trial type in the intentional conditions depended on the participant's decision. The number of trials of each type could therefore not be fixed. Instead, the experiment continued until the participants had done at least 50 trials per condition, or until at least one condition had reached 80 trials. Trials were pseudorandomized, and were in the following proportions: 20% instructed go, 20% instructed nogo, 40% intentional go/nogo and 20% baseline.

Large differences in luminosity have been shown to modulate the perceived duration of stimulus duration (Eagleman, 2008). Therefore, a colorimeter was used to balance the luminosity across colour stimuli.

Before the main experimental session, participants were familiarized with the task in a short practice session. Accuracy feedback was given for the temporal judgement during the practice session. No feedback was given during the experimental session.

Eye movements monitoring

Eye movements were recorded to control for their effect on time perception. In experiment 1, vertical and horizontal electrooculograms (VEOG and HEOG) were obtained using electrocardiography electrodes placed around the eyes of 10 of the 20 participants. The bipolar analogue EOG signal from each channel was amplified (Contact Precision Instruments, London, UK) and collected by a second computer with a sampling rate of 100 Hz, using a 12 bit A/D converter (NI-USB 6008, National Instruments, Austin, TX). The high and low pass filters were set to 0.3Hz and 30 Hz respectively, with a gain of 200 μV. The amplified EOG signal was calibrated immediately before the main experiment. Participants were asked to make large saccadic eye movements from one side of the screen to another(angular size, 31º). These large saccadic eye movements served as a template to determine whether significant eye movements had occurred within the occlusion period. Because saccades can lead to biases in time estimation, it was important to ensure that no differences in relative number of eye movements were present between conditions.

2. 2. 2 Results

Some participants made judgement errors in trials with extreme angular offsets in at least one condition. In these cases (9 participants), not all conditions could be associated with a psychometric curve as a function of the angular offset of the peripheral cue. The indirect temporal judgement task was not easy for all participants, and a high rate of errors in trials with extreme angular offset values was not uncommon. Errors in these trials were the main reason of a failure in the estimation of a psychometric curve. These data were excluded without further analysis, yielding a total of 11 participants.

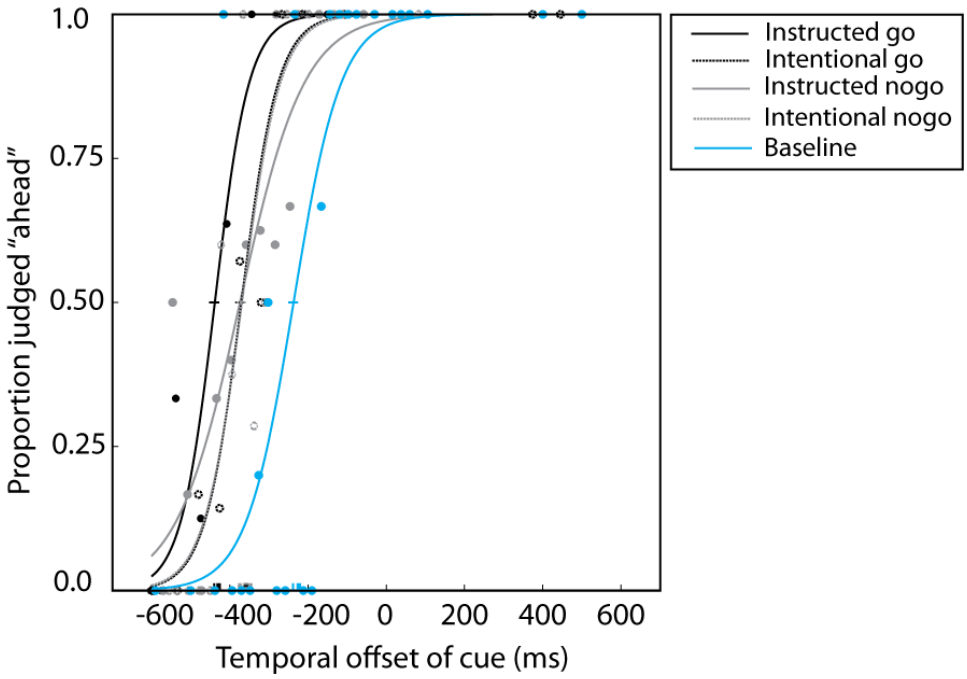
Go/nogo results

Participants made few errors in the instructed trials. The mean (± SD) percentage of omission errors was 1.53 ± 2.02% in the instructed go condition. The percentages of commission errors were 5.18 ± 6.76% in the instructed nogo condition and 0.28 ± 0.63% in the baseline condition. In the intentional conditions, participants successfully produced a balanced outcome between go and nogo, with a mean percentage of intentional go trials of 49 ± 7 %.

A t-test comparing mean reaction times (RTs) between the instructed go and intentional go conditions revealed no significant differences (instructed go, 531 ± 24 ms; intentional go, 548 ± 34 ms; t_{10}=-1.07, p=0309).

Temporal judgement results

To measure the effect of intentional decisions on time perception, the individual points of subjective equality (PSEs) were calculated for each condition. Example data from one representative participant is shown in figure 2.3.



Figue 2.3 Temporal judgement data from a representative participant. The individual data and the fit of the cumulative binomial distribution are shown. Negative shifts of the estimated psychometric curve imply time compression.

PSEs in the baseline condition were significantly shifted towards negative values (-200 ± 120 ms, t_{10} = -5.53, $p<0.001$), indicating time compression. To evaluate the effect of action and inhibition on temporal perception, the mean baseline PSE was subtracted from the mean estimated PSEs for each condition. Baseline-corrected estimated PSEs are shown in figure 2.4.

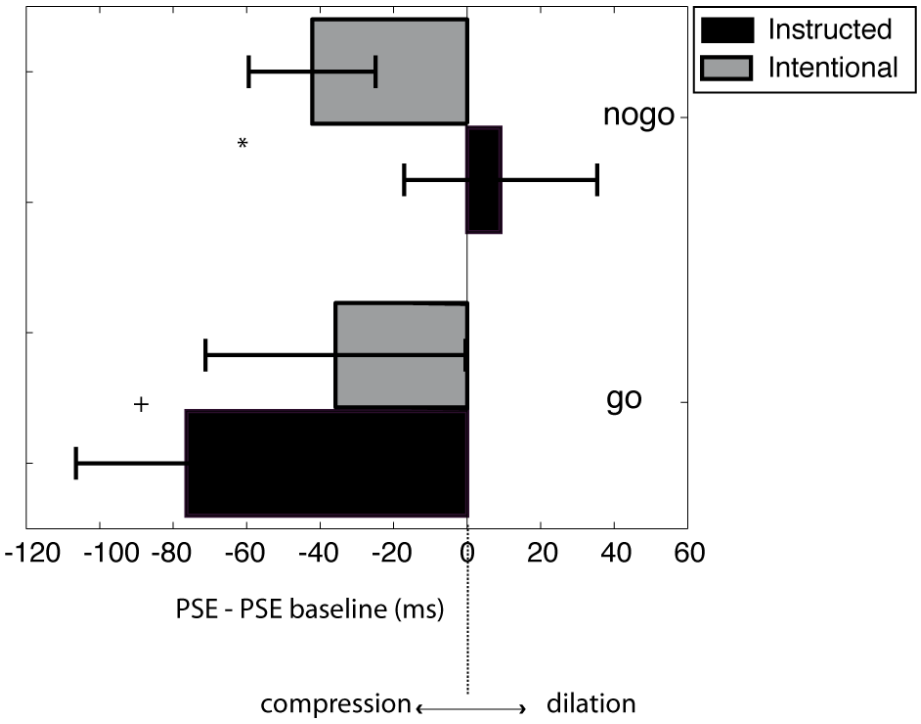


Figure 2.4 Mean points of subjective equality (PSE) relative to baseline PSE, for each condition. Negative PSE values indicate time compression. Error bars show standard error of the mean. * indicates $p<0.05$, + indicates $p<0.1$

A 2x2 ANOVA on the baseline-corrected PSEs with the factors of source (instructed/intentional) and outcome (go/nogo) revealed no main effect of source ($F_{1,10}=0.10$, $p>0.758$) or outcome ($F_{1,10}=2.76$, $p=0.127$); but a significant interaction effect ($F_{1,10}=11.18$, $p=0.007$). Follow up t-tests revealed that the temporal compression associated with instructed inhibition differed from that associated with intentional inhibition ($t_{10}=2.24$, $p=0.048$), with stronger temporal compression in the instructed inhibition condition. There was a non-significant trend for a difference in the opposite direction between the temporal compression associated with instructed and intentional action ($t_{10}=-1.96$, $p=0.078$).

In addition, instructed go and intentional nogo conditions were significantly different from baseline ($t_{10}=-2.5365$, $p=0.029$ and $t_{10}=-2.44$, $p=0.034$ respectively, uncorrected for multiple comparisons). Intentional go and instructed nogo did not differ significantly from baseline ($t_{10}=-1.01$, $p=0.335$ and $t_{10}=0.346$, $p=0.736$, respectively)

Temporal judgement precision

To address whether the go/nogo task had an effect on precision of time perception, JNDs were analyzed in a 2 x 2 ANOVA with the factors of source and outcome. No significant effects were found. There was no main effect of source ($F_{1,10}$=0.78, p=0.399), no main effect of outcome ($F_{1,10}$=0.70, p=0.426) and no interaction effect ($F_{1,10}$=0.22, p=0.648). This suggests that the results cannot be immediately attributed to differences in precision of time estimation.

Eye movements

To control for eye movements, EOG activity was recorded. Eye movements were analyzed offline, after the adaptive QUEST algorithm had yielded the estimated PSEs for each condition. Therefore, the rejection of single trials on the basis of saccades would not be consistent with the adaptive PSE estimation method. A visual inspection of the EOG data showed that saccadic eye movements were rare, with participants typically producing no more than 20 saccades occurring in the total number of trials in the experiment (a minimum of 250). However, the EOG data suggested that despite clear instructions to maintain fixation, participants made smooth pursuits of the peripheral cue in virtually all trials.

2. 2. 3 Discussion

The results of experiment 1 revealed differences in the degrees of time compression associated with intentional and instructed decisions. There were statistically significant differences between instructed and intentional nogo decisions, and a trend for significance between instructed and intentional go conditions. To the extent that biases in time compression are indicative of covert cognitive processes, these results suggest that intentional and instructed inhibition conditions do in fact rely on different sets of mechanisms, since they have dissociable effects on time perception. Whereas instructed inhibition did not differ from baseline in the amount of time

compression, intentional inhibition did show time compression effects over and above baseline.

However, experiment 1 is not informative about the *specific* processes leading to time compression. Biases in time perception can occur at different points relative to action. The time compression effects described by Morrone *et al* (2005) occur in the period around the time of action, whereas intentional binding and saccade-related chronostasis are found in the period immediately following action. In experiment 1, several processes occurred within the occlusion period, and could therefore be responsible for the observed time compression effects. In principle, both pre-action processes (i.e., instruction processing and intentional decision making) and post-action processes (e.g., processes associated with the sensory processing of action consequences) could lead to the observed effects.

2. 3 Experiment 2

2. 3. 1 Introduction

To distinguish the effects of pre-and post-response processing on time perception, experiment 2 aimed at generating a situation of minimal overlap between pre-response processing and occlusion period, by making responses fall just before or early within the occlusion period. The assumption was that time perception biases are primarily due to cognitive processes occurring simultaneously with the period to estimate. Therefore, any time compression effects observed in experiment 2 would be primarily due to post-response processing.

2. 3. 2 Methods

Participants

Fourteen naïve participants took part in experiment 2 (mean age ± SD 26.5 ± 7 years). No participant had taken part in experiment 1.

Task

In experiment 2, participants performed the go/nogo task with their left hand. The task in experiment 2 was as in experiment 1, but with the critical difference that the go/nogo instruction colour change occurred *before* the occlusion period offset (see figures 2.5 and 2.6). In this way, the temporal relation between the to-be-judged period and the response itself was better controlled than in experiment 1. In experiment 1, time compression occurred in the period during action selection and preparation. In contrast, in experiment 2, time compression occurred after action selection, and during action preparation and execution. The decision process will have therefore occurred before the occlusion period. Together, experiments 1 and 2 identified the cognitive processes responsible for the time modulation effects.

The exact time of go/nogo instruction onset relative to the occlusion period onset was estimated for each participant in the first experimental block, and kept constant throughout the experiment.

As in experiment 1, only extreme angular displacements were shown in the first 80 trials. Each participant's mean go RT was estimated on the basis of these initial trials, and used as the temporal difference between go/nogo instruction and occlusion onset. It was assumed that participants with mean RTs longer than 600 ms could improve throughout the experiment. Because the temporal difference between the two task onsets was kept constant throughout the experiment, there would be a risk of overestimating the go RTs due to an initial lack of task familiarity. Therefore, mean go RTs were not allowed to go above 600 ms. That is, each participant's go RT was used as a temporal difference between the tasks, unless it was longer than 600 ms, in which case this maximum value was used.

This method of approximation to the mean go RT was preferred over a strict strategy of initiating the occlusion period upon each action for two reasons. First, to avoid potentially confounding effects of agency related to temporal contingencies between actions and occlusion onset (e.g., intentional binding). Second, and importantly, go RTs could be measured but nogo RTs could not. Because nogo RTs were assumed to be roughly equivalent to go RTs, some error in the temporal matching between the go RTs and the occlusion period made go and nogo conditions more comparable.

In this way, the occlusion period would fall after or around the estimated time of action. Any go/nogo effects on time perception in experiment 2 would therefore be caused by the response itself or by post-response processing, but *not* by decision making or action preparation processes.

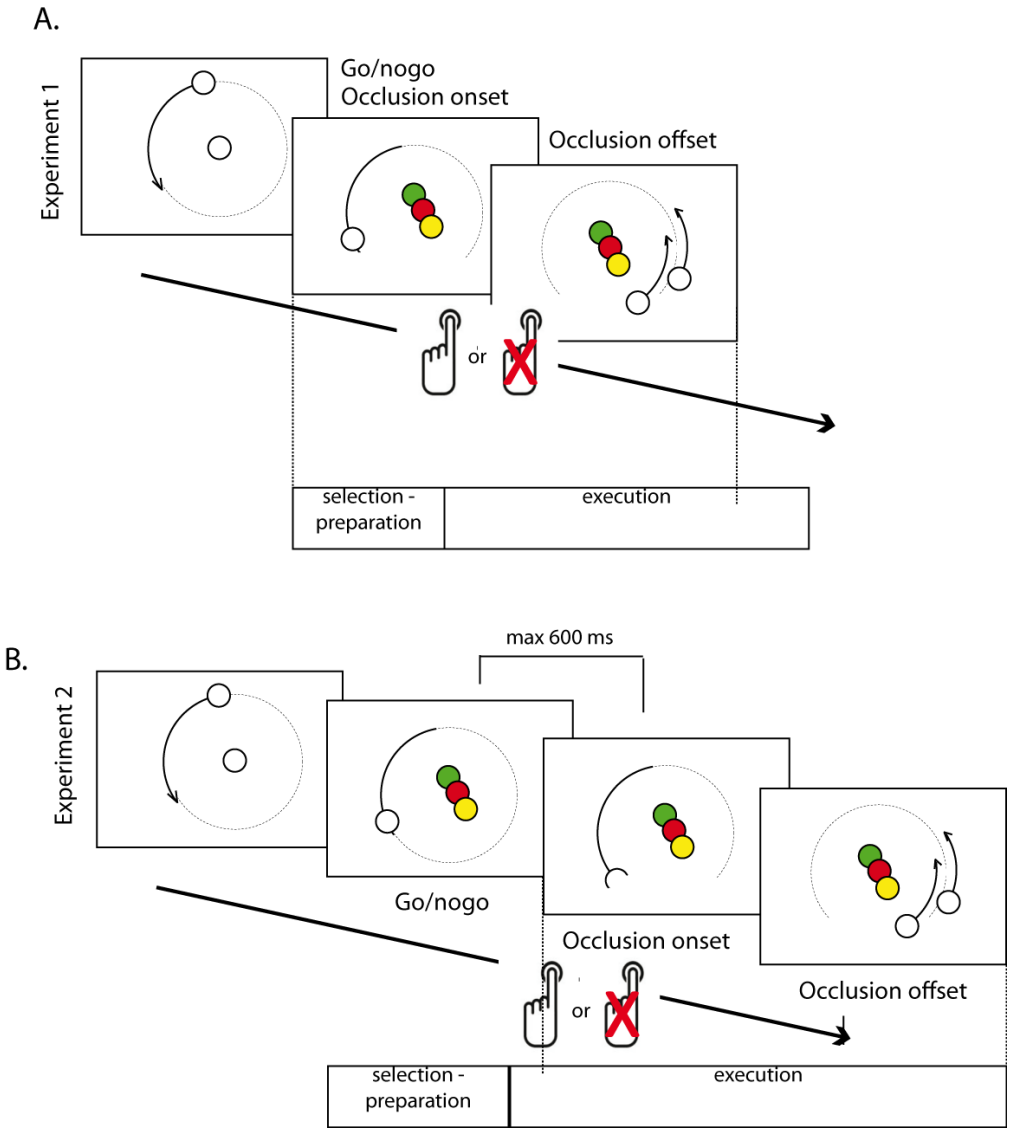


Figure 2.5 Comparison between experimental tasks for experiments 1 and 2. **A,B.** The key difference between experiment 1 and 2 was the onset asynchrony between the time of go/nogo instruction (central cue colour change) and time of occlusion period onset (peripheral cue occlusion). In experiment 1, the two events occurred simultaneously. In experiment 2, the two events were separated in time, with the instruction appearing at a maximum of 600 ms before the occlusion onset (see methods).The mean relative time of the actions with respect to the occlusion period was earlier than in experiment 1.

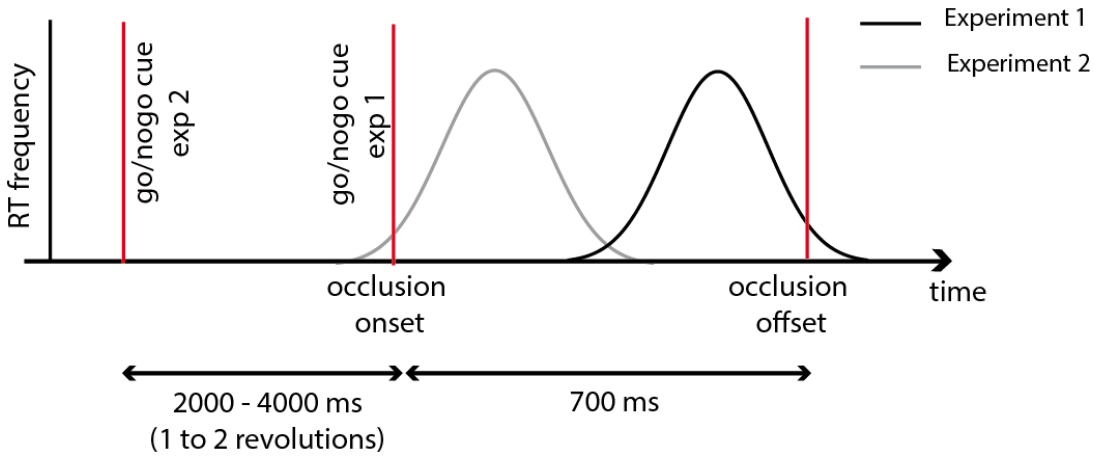


Figure 2.6 Expected probability distributions of RTs in both experiments. In experiment 1, the majority of the RTs fell late in the occlusion period.

Eye tracking

In experiment 2, gaze position was monitored with an eye tracker (Cambridge Research Systems, UK) in 10 out of the 14 participants. Analysis of gaze position was done offline. To determine whether fixation had been broken in each trial, the gaze distance to fixation was calculated for duration of the trial. If it exceeded the radial distance of the peripheral cue (1.5°), fixation was considered to be broken.

2. 3. 3 Results

A psychometric curve for the expected temporal offset of the peripheral cue could not be resolved from the data from 2 participants. These data were excluded without further analysis, yielding a total of 12 participants.

Go/nogo results

Fewer errors were made in experiment 2 than in experiment 1, with a rate of omission errors of 0.02 ± 0.02%, commission errors 0.01 ± 0.02% in instructed nogo condition and no errors in the baseline condition. In the intentional conditions, participants successfully produced a balanced outcome, with a mean percentage of intentional go trials of 51 ± 6 %.

Unlike in experiment 1, mean instructed go RTs were significantly shorter than mean intentional go RTs in experiment 2 (instructed go 630 ± 30 ms, intentional go 717 ± 35 ms, t_{10}=-4.18, p=0.002). As expected, experiment 1 and 2 differed in the time of key press relative to occlusion onset. Relative RTs in experiment 1 were 531 ± 8 ms in the instructed go condition and 548 ± 11 ms in the intentional go condition. In turn, in experiment 2 the mean relative RT was 50 ± 78 ms for the instructed go condition and 128 ± 104 ms for the intentional go conditions. A distribution of the RTs in each experiment pooled for all participants is shown in figure 2.7.

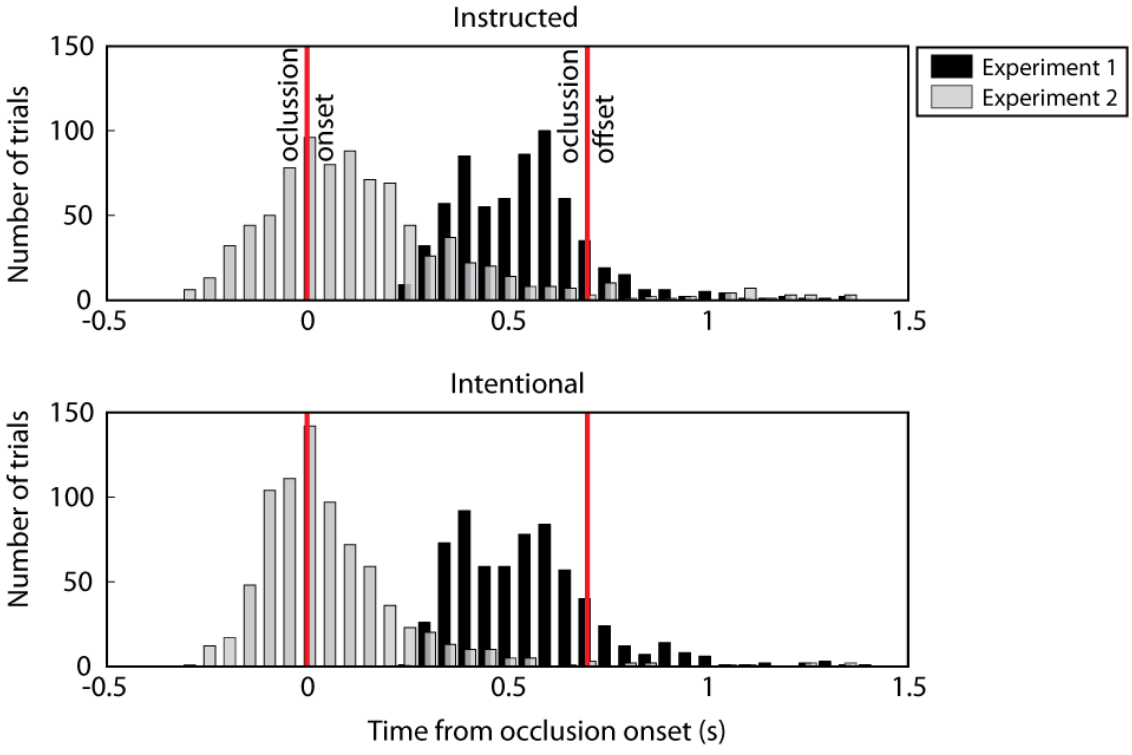
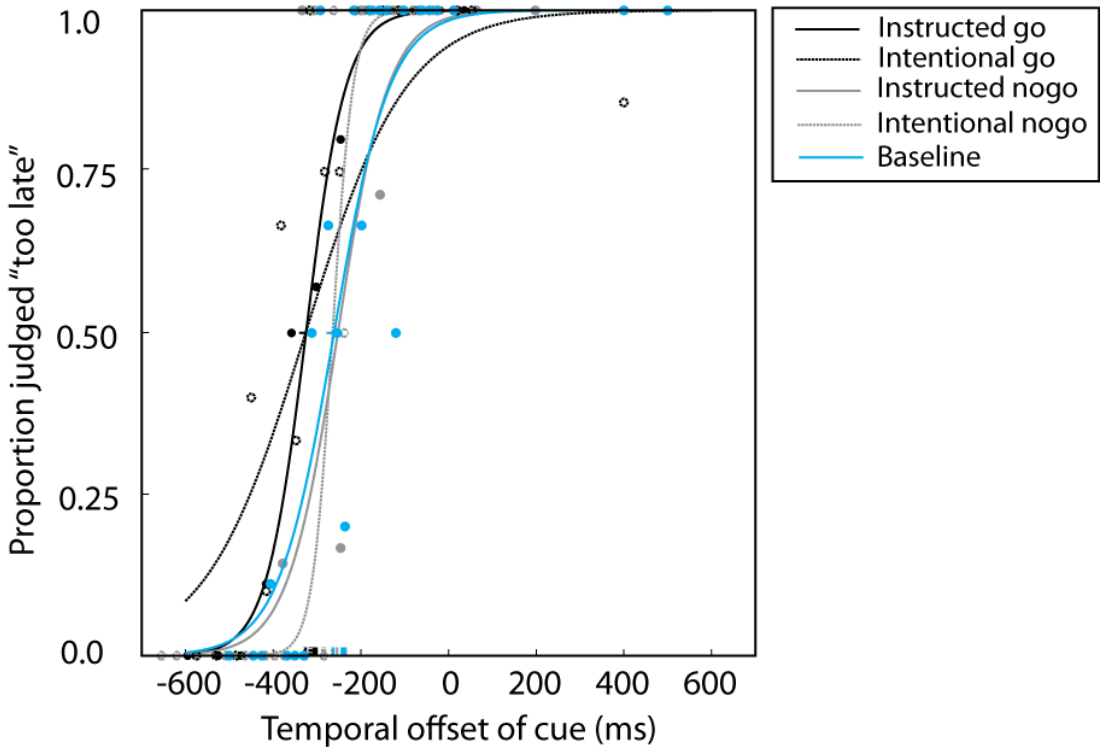


Figure 2.7 Distribution of reaction times relative to the occlusion period. Data are pooled for all participants for experiment 1 (black) and experiment 2 (grey). Red vertical lines mark the onset and offset of the occlusion period. Most key presses occurred late within the occlusion offset in experiment 1, and early within the occlusion period in experiment 2.

Temporal judgement results

As in experiment 1, the effect of intentional decisions on time perception was measured as the PSEs estimated for each condition. Example data from one representative participant is shown in figure 2.8



Figue 2.8 Temporal judgement data from a representative participant. The individual data and the fit of the cumulative binomial distribution are shown. Negative shifts of the estimated psychometric curve imply time compression.

The ANOVA for experiment 2 showed no main effect of source ($F_{1,10}$=1.51, p=0.245) or outcome ($F_{1,10}$<0.01, p>0.99) and no significant interaction ($F_{1,10}$=0.89, p=0.367). No condition showed PSEs significantly different from baseline (all p>0.1440). Baseline-corrected mean estimated PSEs for experiments 2 are shown in figure 2.9.

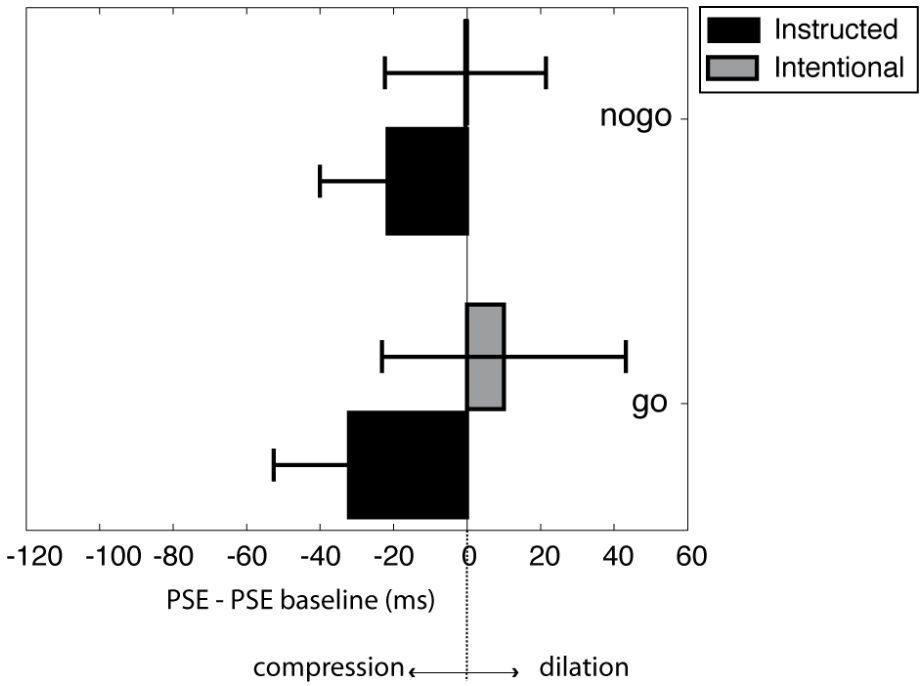


Figure 2.9 Mean points of subjective equality (PSE) relative to baseline PSE, for each condition. Negative PSE values indicate time compression. Error bars show standard error of the mean.

To determine whether the significant interaction in time compression effects was specific to experiment 1, the results between experiments 1 and 2 were compared in a mixed effects 2 x 2 x 2 ANOVA. The baseline corrected mean PSEs for each condition and each experiment were compared with the factors of outcome (go/nogo) and source (instructed/intentional) as within-participants factors and experiment (1/2) as a between- participants factor.

Results of the ANOVA revealed no main effect of source ($F_{1,21}$=0.70, p=0.412) or outcome ($F_{1,21}$=1.40, p=0.249) and no main effect of experiment ($F_{1,21}$=0.84, p=0.369). There was a significant source x outcome interaction effect ($F_{1,21}$=10.32, p=0.004). Importantly, there was a marginally significant three-way source x outcome x experiment ($F_{1,21}$=4.06, p=0.057).

Temporal judgement precision

To address whether the go/nogo task had an effect on precision of time perception, JNDs were analyzed in a 2 x 2 ANOVA with the factors of source and outcome. No significant effects were found. There was no main effect of source ($F_{1,11}$=0.56, p=0.470), no main effect of outcome ($F_{1,11}$=1.13, p=0.310) and no interaction effect ($F_{1,11}$<0.01, p=0.994). As in experiment 1, this suggests that the results cannot be immediately attributed to differences in precision of time estimation.

Baseline comparison

The baseline time compression effect was subtracted from the estimated PSEs for each experiment. The baseline effect was therefore not compared across experiments in the above analysis. A strong mean baseline time compression effect was found in experiment 1 (-200 ± 120 ms, p<0.001). In contrast, no significant baseline time compression effect was found in experiment 2 (-5 ± 20 ms; t_{11} = -1.01, p=0.336). To compare the baseline levels of time compression between the two experiments, a two-sample t-test was done between the estimated baseline PSEs. These values were significantly different (t_{21}=- 2.10, p=0.048)

Eye movements

To control for eye movements, EOG gaze position was recorded with an infrared eyetracker. As in experiment 1, eye movements were analyzed offline, after the adaptive QUEST algorithm had yielded the estimated PSEs for each condition, and the rejection of single trials on the basis of saccades would not be consistent with the PSE estimation method. Offline analyses showed that all participants made pursuit eye movements, in the great majority of trials (68% ± 35%). Therefore, despite explicit instructions to maintain fixation, participants followed the peripheral cue. To determine whether participants made more eye movements in any given condition, the proportion of broken fixations was calculated as the number of trials showing eye

movements for each condition relative to the total number of trials for each participant. A one-way ANOVA with the proportion of broken fixation showed no differences between any of the five conditions ($F_{4,9}$=0.86, p=0.39, Greenhouse-Geiser corrected). Therefore, eye movements are unlikely to explain differences in time perception biases between conditions.

2. 4 Discussion

This study was designed to investigate potential differences between covert cognitive processes associated with intentional and instructed decisions by measuring the consequences for the experienced passage of time. These differences cannot be directly assessed by their overt behavioural manifestations. One common approach in the existing literature is to measure them through their neurophysiological correlates. An alternative approach explored here is to measure covert processes *indirectly*, through their downstream effects on other aspects of cognition, such as through their effects on subjective experience. Here, differences between intentional and instructed behaviour were studied through their effects on subjective time perception. Time perception is widely used as a measure of cognitive processing, and several cognitive aspects of action control have traditionally been studied via time perception (Hagura et al., 2012; Wundt, 1909). Situations that led to identical behavioural outcomes (i.e., either a keypress or no movement) were compared, to reveal differences in their underlying processes.

Experiment 1 measured the effects of decisions whether to execute a prepared action on time perception. Moreover, despite showing virtually identical overt behaviour, measures of subjective experience revealed that intentional and instructed decisions lead to different magnitudes of time perception biases. Because experiment 1 confounded the effects of decision making and response execution, experiment 2 was aimed at discriminating the effects of these two processes. Results suggested that the time compression effects observed in experiment 1 are related to intentional processes that take place before, and not after the execution of a response.

Experiment 1

In experiment 1 all experimental conditions (all except baseline) required participants to make a perceptual decision about the colour change of the central cue during the occlusion period. In contrast, the baseline condition was indicated from the trial onset, and well before the occlusion period. Consequently, in this condition the subjective judgements of the duration of the occlusion period would not have been affected by perceptual judgements of the colour cue, and would provide a measure of "pure" biases in subjective time perception, not related to intentional decisions, action, or inhibition.

In addition, go conditions required action preparation and execution. Assuming that participants were in fact preparing to act from the onset of all but baseline trials, successful nogo trials would have required active action inhibition processes. In this way, nogo conditions would entail action preparation and inhibition. Furthermore, intentional go and nogo conditions required that participants made an additional *whether* decision, to decide on a trial-by-trial basis whether to make an action or not.

Instructed action preparation and execution led to strong time compression effects, over and above baseline. This observation is well in line with the strong time compression effects associated with saccadic movements reported by Morrone et al (Morrone et al., 2005). Intentional action, on the other hand, led to time compression effects that were numerically lower than those for instructed action; there was a trend for statistical differences between these two conditions.

Instructed nogo conditions did not differ from baseline levels of time compression. This may be interpreted in two non-mutually exclusive ways. First, the processes required for instructed nogo trials (namely action preparation and inhibition) may not lead to measurable time compression effects. Second, instructed nogo and baseline trials may have been effectively equivalent. Participants were asked to prepare to act at the time of trial onset in all conditions but baseline. However, action preparation was not otherwise encouraged. If participants were not preparing to make an action, instructed nogo trials would only differ from baseline in that a perceptual decision would have been necessary during the occlusion period in instructed nogo trials.

Time compression in intentional conditions is "intermediate" relative to instructed extremes

These results cannot be easily explained by the filled interval illusion (E. Thomas & Brown, 1974). The filled interval illusion occurs when a time period to be estimated contains a series of perceptual or cognitive events. In these cases, "filled" time intervals are typically perceived as longer than "empty" time intervals of the same objective duration. Intentional decision making could be an additional cognitive process that acts as a "filling". If this were true, time intervals during which intentional decisions took place should be perceived as relatively longer than those where no intentional decision is required. This was not the case. Although the pattern of time compression in go conditions is compatible with this effect, results from intentional conditions are not. Intentional nogo conditions showed *larger*, and not smaller, time compression effects than the instructed nogo conditions. Therefore, intentional processes did not lead to a general positive bias in time estimation, as it might have been expected from results from filled interval tasks. Instead, intentional conditions seem to show "intermediate" effects that lie between the two extremes of the instructed conditions.

No differences between the JNDs were observed across conditions, suggesting that the effects cannot be easily explained by differences in perceptual precision.

There were clear differences in the magnitude of time compression between instructed go and nogo conditions. Strikingly, despite clear differences in behaviour between intentional go and nogo conditions (the former involves overt action, the latter does not), there were no significant differences between the degrees of time compression in the two intentional conditions. Therefore what seems to matter for time perception is not whether movement actually occurs or not, but whether there is action intention and preparation, even if these prepared action plans may be subsequently cancelled. The results of this study cannot be explained by simple motor effects or the filled interval illusion alone. Although both effects may influence time estimation, other cognitive processes in the intentional nogo conditions are necessary to explain the observed pattern of results.

Experiment 2

Results from experiment 1 suggest that the time compression effects observed are related to intentional decision making processes. Two scenarios are thus possible. Time compression effects may be related to decision making and response preparation *per se* (pre-response execution), or they may be related to attributional or compensatory processes following response execution (post-response execution). Experiment 2 was designed to disambiguate between these two possibilities. In experiment 2, the go/nogo cue was given before the onset of the occlusion period. Pre-execution processes such as response selection and preparation would occur immediately after the go/nogo cue, and therefore mostly *before* the occlusion period. In contrast, post-execution processes would mostly occur *during* the occlusion period. Timing of events was approximate, because whereas go RTs could be clearly measured, nogo RTs could only be estimated. It was assumed that cognitive processes will show the strongest effects on the temporal estimation of intervals that occur simultaneously with the mental process. In this way, if the time compression effects observed in experiment 1 were mainly driven by pre-execution processes, separating them in time from the occlusion period would reduce the measured biases in the indirect temporal judgement.

Indeed, the results showed no evidence of modulation of time perception when the pre-response period was temporally dissociated from the occlusion period. A comparison between experiments 1 and 2 revealed marginally significant differences, suggesting that the effects observed in experiment 1 may in fact be specific to the pre-response period. Therefore, the temporal compression observed here must be related to response preparation and selection, not response execution

Differences in baseline effects of time compression across experiments 1 and 2

In experiment 1, a strong time compression effect was observed in the baseline condition. This effect was greatly reduced in experiment 2. Baseline trials were identical across both experiments, so the reasons for this strong difference are not easy to understand. However, an analysis of the mean go RTs and error rates shows strong differences in go/nogo task performance between the two experiments. Go

RTs were quicker in experiment 1, and error rates were higher than in experiment 2. This suggests that participants in experiment 1 prioritized speed over accuracy in the go/nogo task. In experiment 2 participants may have taken more time to do the task, and may have allowed for better accuracy in the time duration judgement task.

Mechanisms of time perception

The neural mechanisms responsible for time perception are still a matter of debate. Two basic opposing mechanisms have been proposed (Ivry & Schlerf, 2008). First, time perception may rely on specialized mechanisms or neural structures. Under an alternative prevailing view, the representation of interval durations may rely on certain inherent properties of all neural systems. Thus, time perception is ubiquitous and depends on the perceptual modality related to the timing task (Eagleman, 2008). Further, it has recently been suggested that subjective duration of a perceptual stimulus is related to the amount of neural firing used to encode a stimulus (Pariyadath & Eagleman, 2007). Crucially, this implies that the perception of time does not only depend on the neural structure related to the specific perceptual modality, but also depends on the current state of the neural network. For example, Hagura *et al* (2012) have shown that the magnitude of action-related biases in time estimation depends on the amount of action preparation. The effects reported by Hagura *et al* are related to time dilation, and not the time compression effect observed here. However, it remains a speculative possibility that action preparation in the context of the go/nogo task also modulated the observed time compression effects.

In humans, BOLD signal increases have been found in parietal, motor and supplementary motor areas (inferior parietal lobe -IPL-, dorsal premotor cortex -PMd- and supplementary motor area -SMA-) in tasks requiring duration judgements of intervals in ranges of both second and subsecond scales (e.g., Lewis & Miall, 2003; Schubotz, Friederici, & Yves von Cramon, 2000). This result is in line with data from monkey electrophysiology, showing links between both sensorimotor

parietal areas and motor areas in time perception (Janssen & Shadlen, 2005; Leon & Shadlen, 2003; Roux, Coulmance, & Riehle, 2003).

Together, these results suggest that parietal and motor areas, typically related to sensorimotor integration and motor preparation, are also involved in mechanisms of temporal estimation, providing a tight link between the two functions. In particular, if premotor and motor cortices are involved in mechanisms of temporal estimation in this task, then their underlying state of motor preparatory activity will have an effect on time perception. In the absence of external imperatives, intentional conditions may lead to intermediate degrees of motor preparation because no clear and unambiguous instruction is provided. The generation of responses will not be clear cut, and the two response alternatives (in this case, go and nogo), may both be partially activated (Cisek & Kalaska, 2005). In contrast, in the instructed conditions, clear external instructions may lead to an efficient suppression of the "incorrect" response. Therefore, in light of these findings, a possible explanation for the results observed here is that the degree of time compression is related to the *strength* of action preparation processes.

Conditions with strong motor preparation (such as instructed go) may in turn lead to strong biases in temporal estimation. Conditions with low levels of action preparation (in this case, instructed nogo) would lead to minimal biases in temporal estimation, as are comparable to baseline levels. Crucially, if intentional conditions - both intentional go and intentional nogo- represent levels of action preparation that are intermediate between the instructed go and nogo conditions, then, these conditions would show intermediate levels of biases in time compression.

This speculation is supported by data from monkey electrophysiology. Watanabe et al (Watanabe, Igaki, & Funahashi, 2006) trained monkeys in an oculomor delayed response (ODR) task. In a canonical ODR condition, analogous to the instructed conditions considered here, monkeys were shown one target, to which they had to saccade after a 3 s delay. In a self-ODR condition (S-ODR), analogous to the intentional conditions discussed here, monkeys were shown four different possible targets, and they would have to saccade to either of them after the same 3 s delay. Watanabe *et al* recorded neural activity from neurons in the dorsolateral PFC (DLPFC). They found that neural firing in the ODR task was high and sustained

during the delay period for the instructed direction only. In contrast, neural firing during the delay period in the S-ODR condition was not directional, and was generally lower than that in the instructed conditions. Neural firing in S-ODR conditions became directional only at the end of the delay period.

These results suggest that intentional conditions are associated with "weaker" neural codes than otherwise behaviourally identical instructed conditions. These partially activated neural codes were not only present in motor-related areas; Watanabe *et al*'s results are drawn from neural firing in DLFPC. Given that time perception seems not to depend on a specialized structure, but instead be subserved by distributed neural networks, it is possible that weaker motor-related neural codes in cases of intentional action lead to the "intermediate" levels of time compression.

The results described by Watanabe *et al* are related to the neural representation of decisions about *what* action to make. There is of course no direct evidence that the same "partial representation" mechanisms occur also for early and/or late decisions about *whether* or not to make an action. It is however plausible that when an intentional decision about whether to act or not is required, weaker (or intermediate) motor preparation occurs. In other word, the suppression of motor preparatory activity, consistent with a nogo response, may not be as efficient as in instructed cases.

Limitations of this study

One important limitation of this experimental task is it cannot guarantee that action preparation occurred in every experimental trial (i.e., in non-baseline trials). Participants were asked to prepare action at the start of every trial. However, no formal comparison between conditions of action preparation *vs*. no action preparation can be made, and the overall balanced numbers of go and nogo trials were not optimally designed to encourage action preparation. Therefore, the assumption that participants actively inhibited actions that they had already prepared to a late stage should be taken with care. This limitation is important when taking into account the differences between early selective decisions about whether to act or not *vs*. late inhibitory decisions (see the introduction of this thesis, section 2.2). If

action preparation and active inhibition occurred within an instructed nogo trial, then the present results do indeed reflect differences at the level of the late inhibitory decisions. If participants did not prepare actions, these results should be interpreted as related to early selective decisions about whether to act or not, rather than as late inhibitory processes. These early decisions are closer to response selection processes, whereas inhibitory decisions are closer to response inhibition processes. These results cannot be unambiguously related to a given decision stage. Nevertheless, this study allows to clearly compare intentional and instructed conditions, regardless of the timing of the nogo process.

Further, the results from experiment 1 suffered from a high exclusion rate (9 out of 20 participants). Data from individual participants were excluded from further analyses when at least one condition could not be fit by the cumulative binomial function used to estimate the psychometric curves. This high exclusion rate suggests that the task was generally difficult for participants, even at the extreme values of angular displacement from the PSE. These results should therefore be taken with care, and future manipulations should ensure more reliable behaviour around the extreme values of angular or temporal displacement.

Finally, the present results may be confounded by eye movements. To avoid the well-documented effects of eye movements on time perception, participants were instructed to maintain fixation. However, EOG and eye tracking data show that participants followed the revolution of the peripheral cue in the vast majority of the trials. Smooth eye pursuit movements have been shown to lead to time compression (Schütz & Morrone, 2010), which may explain the baseline time compression effects. However, and critically, no differences in eye movements were found between conditions, suggesting that although a systematic time compression effect may be due to eye movements, they cannot easily explain the differences between conditions.

2. 5 Conclusion

Two experiments showed that measuring the subjective experience associated with action and inhibition can reveal differences in covert cognitive processing that leads to a given behavioural outcome. In this case, the biases in time perception associated with actions and inhibition of action depended on whether the decision to act or inhibit had been intentional or instructed. Interestingly, the results suggest that intentional decision-making processes necessary for intentional inhibition can be dissociated from both instructed inhibition and nonaction. In addition, intentional decisions were associated with intermediate levels of time compression that lied between the two extremes of instructed actions and nonacions. Intentional decisions to act or inhibit may be associated with weaker neural codes than instructed decisions. The effects on time experience of intentional inhibition are similar to those of intentional action. In turn, both are intermediate between the compression associated with instructed action and the compression associated with instructed inhibition. Thus experimentally, intentional action and inhibition seem to occupy "intermediate" positions irrespective of whether the action occurs or not. This may reflect aforementioned effects of levels of action preparation.

Chapter 3 Inhibiting the urge to avoid an itch

A naturalistic paradigm for studying intentional inhibition was developed by giving participants the choice to make an action that terminated an itch, or to inhibit the action and tolerate the itch. Event-related potentials evoked by the itchy stimuli were then used to compare the consequences of intentional vs. instructed inhibition of the itch-terminating action. The amplitude of the event-related potentials evoked by itchy stimuli was similar when participants had intentionally decided to act to avoid the itch, or had intentionally decided to inhibit action and tolerate it. In contrast, instructed action and instructed inhibition produced quite different itch-evoked ERPs. In terms of their consequences for later processing, intentional decisions to inhibit appear close to intentional decisions to act, consistent with the concept of a finely-balanced process for either enabling or disabling intentional action.

3. 1 Introduction

Inhibitory self-control

The concept of intentional inhibition partly overlaps with the concept of willpower developed in behavioural social psychology (Baumeister et al. 2007), in that both have the effect of preventing or delaying inappropriate actions. For example, most people can recognize the urge to scratch an itch. *Not* scratching can be extremely effortful, and it can make the itchy feeling more intense. Here an experimentally controlled version of this situation was developed to create a paradigm that would meet these requirements.

Recent experimental work has given rise to the idea that willpower, or self-control, is a general capacity, or limited resource analogous to the body's physical energy (Hagger et al. 2010). The inhibitory mechanisms associated with willpower relate to a general state in which inhibition is continuously present, until exhausted. However, intentional inhibition might also involve a temporally-specific decisional process, analogous to instructed inhibition triggered by a stop signal. For example, it has been argued that people may withhold an intentional action at the last possible moment (Libet 1985). Therefore in an event-related framework for intentional inhibition, inhibition should appear not only as a general, sustained mental process, but also as a specific, clearly-timed event. That is, the decision to inhibit may be taken in the context of a specific stimulus, and may have specific consequences. Treating inhibition as event-related would allow the neural mechanisms to be measured more precisely.

Experimental constraints raise significant methodological problems, notably for ecological validity. Most previous experimental studies of intentional action and inhibition have not given participants clear reasons for choosing to act or inhibit, and thus have low ecological validity (e.g., Libet et al. 1983; Brass and Haggard 2007). This was also the case in the experiments described in chapter 2 of this thesis. The need for methodological simplicity has meant that reasons, urges, values and consequences of actions have been conspicuously missing in these paradigms.

Measuring the consequences of inhibition

One recent experiment suggests that the sensory consequences of action could be useful to describe the processes of action that were in fact inhibited. Shocks delivered to participants' fingers were perceived as weaker after action inhibition triggered by an external stop signal, as compared to a passive detection task (Walsh and Haggard 2010), suggesting that some characteristics of intentional action are maintained even if the action itself is inhibited. An experimental framework to study intentional inhibition of action requires three components. First, there must be a reason to perform an action. Second, the participant must make an intentional decision to inhibit that action on some occasions. Third, there should be some way of measuring the intentional processes associated with inhibition, rather than merely recording whether an action occurred or was inhibited.

Resisting the urge to scratch an itch

In this ecologically valid task, reasons for actions were provided by delivering on each trial itchy and unpleasant stimuli that could be avoided by doing a hand movement. In this way, decisions to act or inhibit would have meaningful consequences. Intentional decisions about action and inhibition were allowed in some trials, whereas clear instructions were provided in other trials. Electroencephalography (EEG) was recorded to measure brain activity.

Specifically, participants were either instructed or had to decide whether to move their arm to avoid an unpleasant itchy sensation, or to inhibit the urge to move the arm, and withstand the itch. Thus, a strong motivation to act was introduced.

Event-related potentials (ERP) were compared for situations of instructed vs. intentional inhibition. Sensory processing of itchy stimuli produced a strong event-related potential (Mochizuki et al. 2008). Therefore, ERPs were measured to take advantage of the good temporal consistency of these known potentials.

The main hypotheses were as follows. First, a decision to execute or inhibit action

will influence subsequent sensory processing. Second, and most importantly, this influence will vary with the source of the decision: intentional decisions would have different 'downstream' effects on sensory processes from instructed decisions. In following with the results described in chapter 2, intentional decisions to execute or inhibit action were expected to show intermediate levels of sensory processing of the stimuli, lying between the extremes of the sensory processing associated with instructed action and inhibition.

3. 2 Experiment 1

3. 2. 1 Methods

Participants

Sixteen naïve paid healthy volunteers (9 females, mean age 25 ± 5 years) participated in the experiment. Participants with sensitive skin were excluded from taking part in the study. Procedures were approved by the UCL research ethics committee and were in accordance with the principles of the Declaration of Helsinki. One participant was excluded after participation due to excessive blinking.

Participants sat at a table and 60 cm away from a computer screen. Both hands rested comfortably on the table, so that their right index finger would rest on a force-sensitive resistor (FSR) (Active Robots Ltd, Somerset, UK), connected to a computer.

Electrical Stimuli

Itchy stimuli were delivered using previously established methods (Mochizuki et al. 2008) with some adaptations. Briefly, each electrode consisted of four pairs of stainless steel wire 0.1 mm in diameter. Each pair formed a cross and was placed approximately 2 mm away from the next pair (see figure 3.1). Current for each electrode was supplied by a Digitimer medical stimulator (Digitimer Ltd, Hertfordshire, England), and flowed through all wires. The reference electrodes

(cathodes) were placed 1 cm laterally to each itch electrode. Current was delivered through the electrodes in square pulses of 2 ms duration at 50 Hz. A stepwise procedure was used to determine the current intensity necessary to elicit an unpleasantly itchy sensation in each participant. The current was explicitly set to a level that would rather be avoided, but would still be bearable. Participants were asked to rest their right hand on a force-sensitive pad. Electrical stimulation would flow through the electrodes on their left forearm only as long as the pad was touched by the right hand. As soon as the participant withdrew their hand from the pad, stimulation would stop and the itchy sensation would cease. Thus, participants could stop an unpleasant sensation on their *left* wrist by actively withdrawing their *right* hand from a resting position. Critically, while participants held their hand in place on the pad they actively inhibited the urge to act.

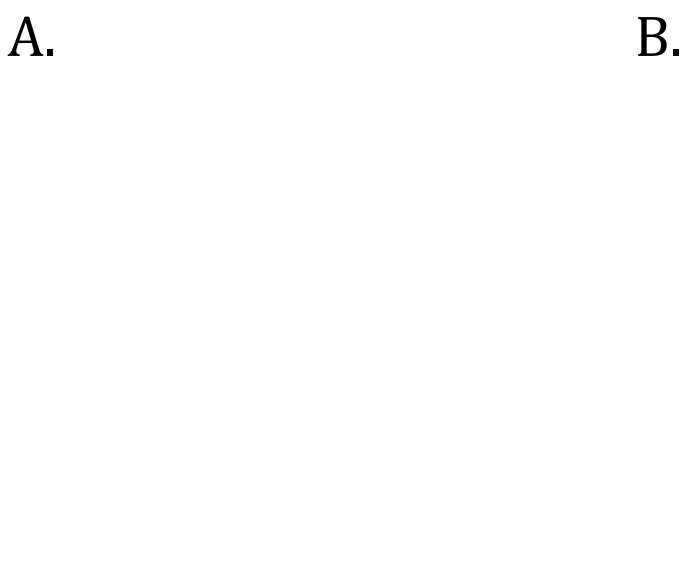


Figure 3.1 A. The itch electrode used for stimulation and **B.** Positioning of the electrode on the participants' wrists. Adapted from Mochizuki et al (2008)[1]

As expected from previous studies (Mochizuki et al. 2009), participants reported a strong habituation of the itchy sensation with repeated stimulation. Therefore, two separate itch electrodes were placed on the left wrist. Stimulation lasted for 3

[1] I thank Ryusuke Kakigi for providing a prototype of the itch electrode.

seconds altogether and consisted of three 1 s-long shocks, alternating from one electrode to the other. The first electrode to be stimulated (proximal or distal to the wrist) alternated across trials.

Task and experimental design

The experiment consisted of 10 blocks of 40 trials each, and lasted 80 to 90 minutes. Each trial was organized as follows: (see figure 3.2) a black fixation cross appeared over a grey background for a variable duration of between 2 and 3 seconds. Two visual stimuli (called V1 and V2) were presented sequentially. These signalled respectively the start of each trial, and the instructions for a given trial. V1 was a green circle subtending 1.5º at a distance of approximately 60 cm, appearing for 250 ms to mark the initiation of the trial. Participants were asked to prepare a right hand movement as soon as V1 appeared. The fixation cross then appeared again on screen for 2 s, until a second circle appeared (V2), again for 250 ms. The luminosity of V1 and V2 was balanced with an heterochromatic flicker test with an independent set of 5 participants. This was done to adjust the intensity of the visual stimuli independently of their absolute luminance, but depending on participants' sensitivity (Wyszecki and Stiles 1982). V2 was of the same size as V1 and could be of three different colours. If V2 was green, participants should remove their right hand as soon as they felt the shock on their left, thus terminating the shock.

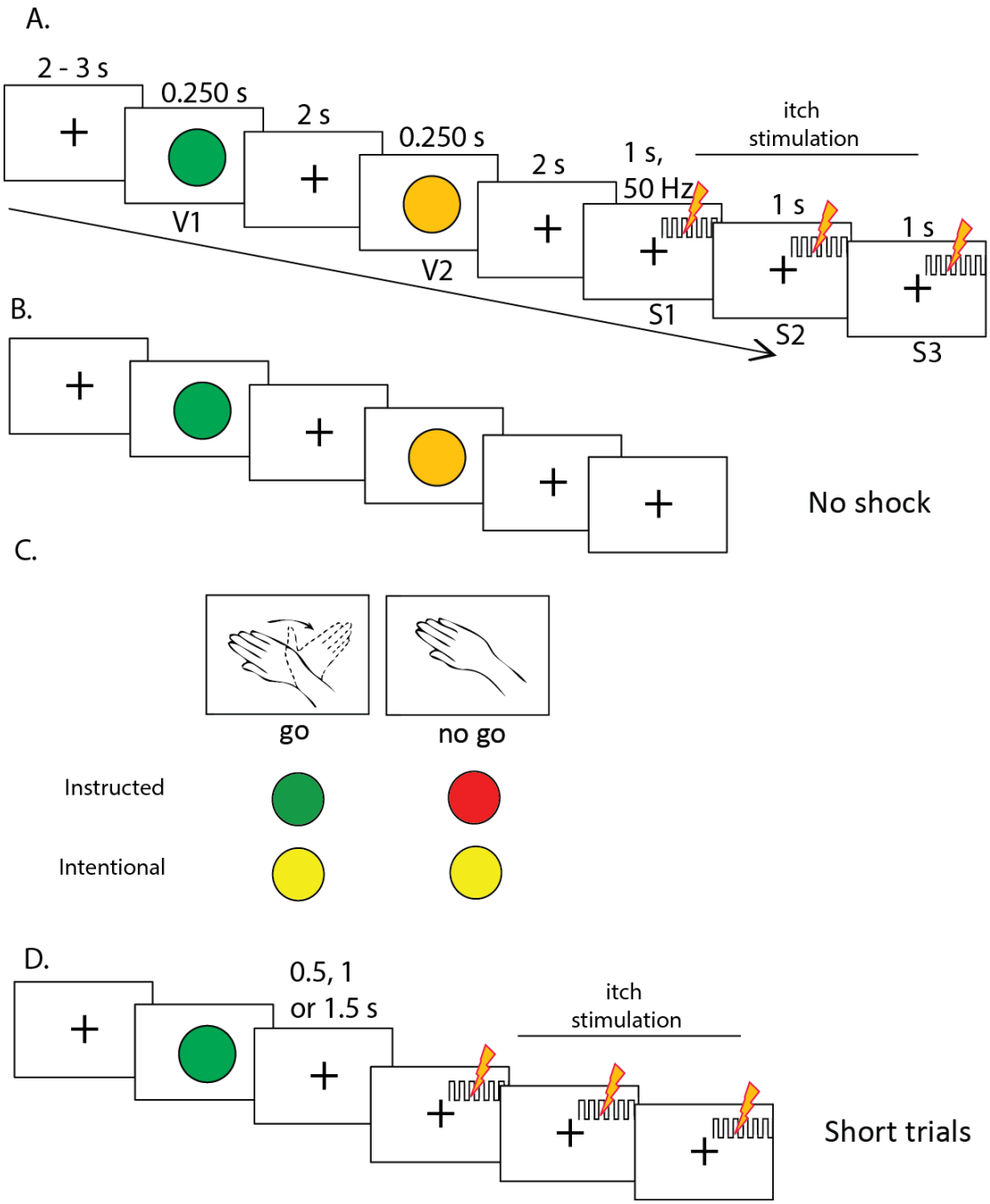


Figure 3.2: Experimental task. **A.** General timing of go and no go trials. A visual warning sign (V1) was presented for 0.250 s. After a 2 s interval an instruction sign (V2) was presented, also for 0.250 s. After a second 2 s interval, three consecutive electrical shocks were delivered at 50 Hz and for a maximum of 3 s, generating an itchy feeling on the participants' wrist. If the participant moved their hand, they would interrupt the itchy feeling. **B.** In no shock (catch) trials, no shocks were delivered to ensure that participants were waiting for the shock to execute their decisions. **C.** Factorial response mapping for trials in A. and B. Green and red V2 corresponded to instructions to move the hand or not, respectively. Yellow V2 allowed the participants themselves to choose between the two possible action outcomes. Short trials in which V2 was replaced at variable times by a surprise shock were presented to encourage and probe motor preparation. Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

The green V2 represented an instructed go condition, in which participants were instructed to perform a movement. If V2 was red, participants were asked to endure the shock, and were not to move their hands. Hence, the red V2 represented the instructed no go condition, in which participants were instructed to avoid doing any movement despite presumably having the desire to do so.

Alternatively, V2 could be yellow, in which case the participants were asked to decide whether they would endure the shock (intentional no go) or withdraw their hand (intentional go) as soon as they felt the shock, thus avoiding prolongation of the itchy sensation. Participants were encouraged to choose to withdraw their hand in roughly 50% of the yellow V2 trials. After each block of 40 trials, they received feedback if the rate of withdrawal was higher than 70% or lower than 30% of the trials. Because participants preferred to withdraw their hands whenever they had the choice, this manipulation ensured that the overall number of trials were comparable across conditions for the EEG analysis. The block-by-block feedback and relatively loose boundaries around 50% were included to prevent participants from developing a very strict strategy.

In this way, the experiment followed a factorial 2x2 experimental design, with the factors of source (intentional/instructed) and outcome (go/no go). Critically, there was no behavioural difference between the intentional and instructed conditions, so any differences found in the recorded neural signal associated with the intentional or instructed sources of decision would necessarily reflect differences in the processing of intentional *vs*. instructed decisions.

Two additional conditions were included for methodological reasons: 15% 'catch' trials (without a shock) were presented to ensure that participants waited for the first shock before executing their intentional decision or the instruction, and did not simply predict its onset. All three V2 colours were followed by catch trials with equal probability. In addition, to encourage movement preparation, 25% of 'short' trials were included. In these trials itch stimulation was delivered at either 0.5, 1.0 or 1.5 s after V1, in contrast with the normal time of 2250 ms. Participants were asked to withdraw their hand from the FSR as quickly as possible in these cases.

The mean intensity at which participants reported to feel an unpleasant but bearable

itchy sensation was 0.36 ± 0.14 mA at the beginning of the experiment for both electrodes. After each block, intensity was readjusted if the stimulation was perceived as too painful or too mild. Intensity never exceeded 0.4 mA, and by the end of the experiment the mean intensity at which subjects perceived the itchy sensation was 0.38 ± 0.14 mA and 0.38 ± 0.15 mA respectively for each one of the stimulators.

Electrophysiological Recordings and signal analysis

A SynAmps amplifier system and Scan 4.3 software (Neuroscan, El Paso, TX) were used to record EEG data. Activity from fourteen scalp electrodes was recorded (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, P3, Pz, P4, O1, O2, according to the 10-20 system). The reference electrode was AFz and the ground electrode was placed on the chin. All electrode impedances were kept below 5 KΩ. The left and right mastoids were recorded. Horizontal electroculogram (EOG) was recorded from bipolar electrodes placed on the outer canthi of each eye, and vertical EOG was recorded from bipolar electrodes placed above and below the right eye. EEG signals were amplified and digitized at 500 Hz.

EEG data were analyzed with EEGLAB software (Delorme and Makeig 2004). Data were first re-referenced to the linked mastoids. Because long epochs (8.25 s) were defined, data were digitally high-pass filtered over 0.5 Hz to remove low frequency drifts. In addition, the amplitude of event related potentials (ERPs) were calculated as peak amplitude values. A 30Hz low-pass filter was applied to the data (Mochizuki et al. 2008). Continuous EEG data was time-locked to the trial start (stimulus V1), and baselined to the period of 250 ms to 150 ms prior to the onset of V1. This baseline fell long before the time of decision (V2). This early baseline was desirable because trials were classified according to the participants' intentional decisions and it was desirable to take a baseline before a decision was made. To avoid artefacts due to eye blinks, trials were discarded if the bipolar recording of EOG exceeded ±80 μV at any point during the epoch. The mean percentage of rejected trials was 22%. This value is relatively high, but perhaps unsurprising given the long epochs and the unpleasantness of the experience. The components in the evoked response

were identified by inspection of the grand average pooling across all conditions. For each component identified in the grand average, the time of maximum amplitude of the individual average was determined and the values for each participant in that time point were computed.

3. 2. 2 Results

Behavioural results

Participants rarely moved their hands in catch trials (mean ± SD commission errors 0.70% ± 0.91%).

On average, within the intentional trials, participants decided to withdraw their hands (intentional go) on 46 ± 5% (mean ± SD) of the trials. The average RT to withdraw the hand after receiving an itchy shock was calculated in order to measure the extent of preparation of the action to withdraw. RTs were compared across the intentional and instructed go conditions (for which the withdrawal movement could be anticipated and prepared) and the average of all 'short' trials, in which the shocks occurred without a prior V2 warning signal, therefore not allowing for movement preparation (3.3).

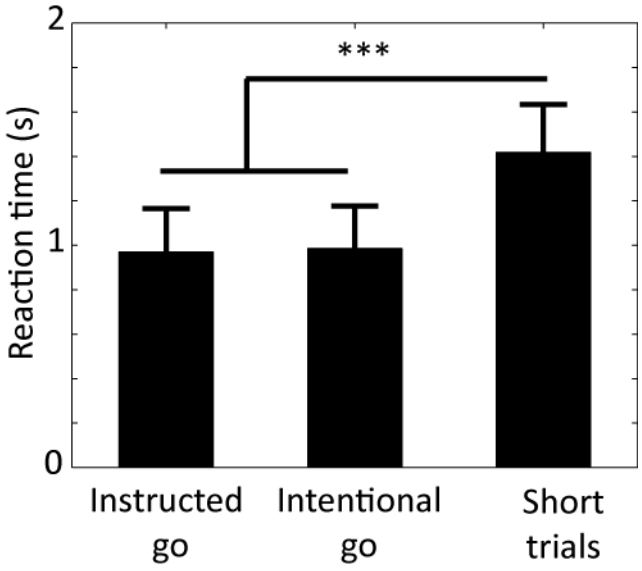


Figure 3.3: Behavioural results. Hand withdrawal times for the movement conditions. No difference is observed between the two movement conditions, in which the time of shock was highly predictable. Longer withdrawal times are observed in 'short' trials, suggesting that motor preparation had occurred in both intentional and instructed *go* trials. Error bars show confidence intervals. *** represent $p<0.001$. Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

A repeated measures ANOVA revealed a main effect of condition ($F_{1,15}=19.38$, $p<0.001$). RTs were longer for short trials as compared to both intentional go and instructed go conditions. Paired comparisons revealed significant differences between the intentional go condition and the short trials ($t_{15}=-6.22$, $p<0.001$) and between the instructed go condition and the short trials ($t_{15}=-5.25$, $p<0.001$). No significant differences were found between the RTs for the intentional and instructed go conditions ($t_{15}=-0.54$, $p=0.59$). These results suggest that there was movement preparation in the two go conditions that was less efficient in the short trials. It is also possible that go trials had shorter RTs due to the timing of the shocks becoming more predictable. Our design cannot distinguish between these two possibilities.

Relatively long RTs were observed (of around 1 s). Such long reaction times may partly reflect the peculiar nature of this stimulation. An interesting feature of this itch stimulus is the lack of a discrete perceptual onset at the start of the shock-train. Short trains do not produce any sensation at all. At the intensities used, sensory perception

began only some time after the onset of the stimulation. Because the reaction time is measured from the onset of the stimulus, the reaction time is artificially increased by delay, which may be attributed to accumulation of signals in perceptual areas.

These behavioural results did not change after excluding the participant that was excluded from the ERP analysis due to excessive blinking.

ERP results

After blink rejection, an average of 42 ± 16 trials (SD) were recorded for the intentional *go* condition, and 41±17 trials were recorded for the intentional *no go* condition. 42±13 trials were recorded for the instructed go condition, and 49±13 trials fell into the instructed *no go* condition. The grand-average ERPs were displayed time-locked to V1, to reveal the sequence of sensorimotor events in each epoch. Figure 3.4 shows the grand-average trace at C3, Cz and C4 pooled across all conditions. There is a stereotyped response to the onset of both V1 and V2. Importantly, although V1 and V2 are physically similar, only V2 carried information about the subsequent task instructions. Accordingly, the neural processing of V1 differed strongly from that of V2, with only V2 eliciting a strong positivity peaking at around 580 ms after V2 onset. There was a characteristic negativity preceding V2, recalling the CNV (Lumsden et al. 1986). This negativity started roughly 800 ms before the onset of the first shock is visible in the grand average across conditions. Finally, the neural response to the three consecutive shocks was apparent. A marked positive-going component occurs in response to each of the three shocks, peaking at around 400ms after shock onset.

The key ERP components evoked by the first two shocks are indicated by shaded areas A and B in Figure 3.4. It shows an average of all conditions, including both go and no go conditions. Therefore, whilst the first shock was always delivered, and acted as a go signal, the second and third shocks were not experienced if participants withdrew their hand. In addition, the third shock was hardly perceived, even on no go trials, due to habituation. It was therefore not included in the analysis. Peak amplitudes for each of these events were analyzed for each condition separately.

The analysis focussed on the effects of source of decision (intentional or instructed) and decision outcome (go or no go) on the neural activity evoked by the shocks. Further, the analyses were restricted to central and parietal electrodes, because the main interest was to assess the neural consequences of inhibition over sensory processing, rather than the frontal mechanisms that cause inhibition itself (Aron et al. 2004).

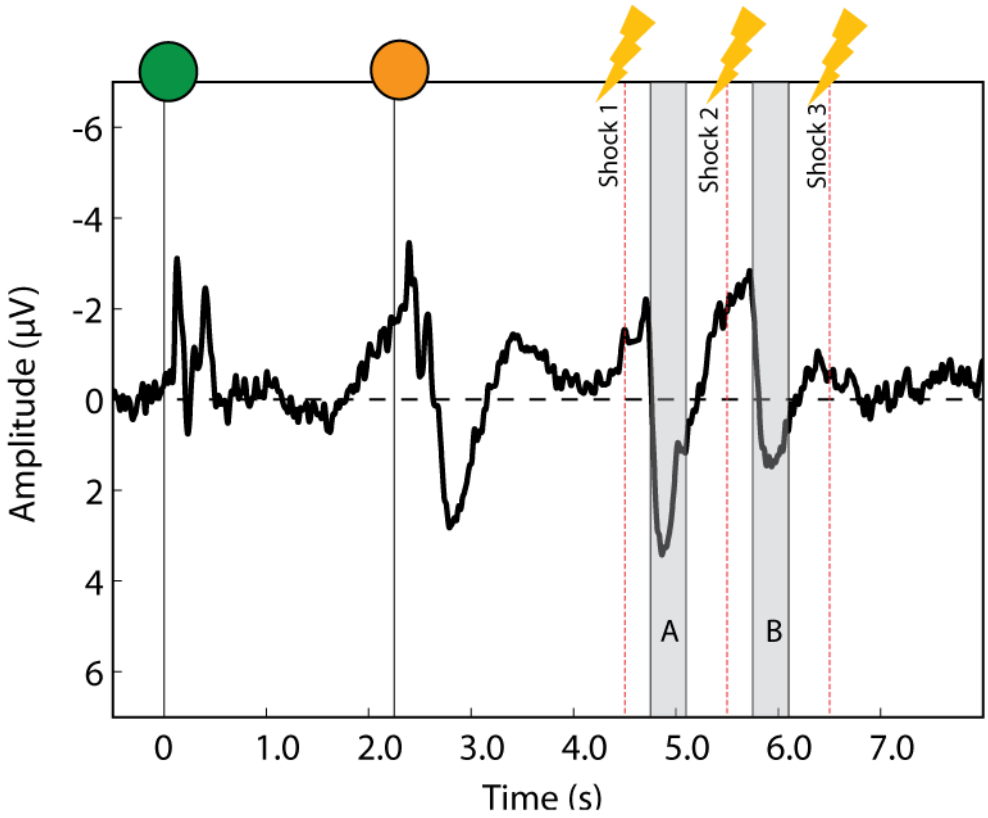


Figure 3.4: EEG results **(a)** Grand average across all the four main experimental conditions (intentional and instructed *go* and *no go*) for electrodes C3, Cz and C4. Long (8 s) epochs were visually inspected to identify two main periods of interest: response to shock 1 (*A*) and response to shock 2 (*B*). The response to shock 3 was greatly attenuated and was therefore not analyzed. EEG trace is plotted negative-up. Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

Pre-shock components

First, the response to the instruction cue was examined. A time window of 350-550 ms after V2 was chosen. In accordance to the topographical distribution in the half-point of the chosen time window, electrodes C3, Cz, C4, P3, Pz and P4 were averaged. A 2x2 ANOVA of the peak amplitudes revealed a trend for a main effect of outcome ($F_{1,14}$=3.42, p=0.08), with a stronger V2 positivity in the no go conditions. There was no main effect of source ($F_{1,14}$=0.39, p=0.53) or interaction effect ($F_{1,14}$=0.03 p =0.59).

Then the preparatory activity the V2 instruction and before shock 1 was examined. An inspection of the grand average (figure 3.4) shows that there is an RP/CNV component before shock 1 (Kornhuber & Deecke, 1965; Walter, Cooper, Aldrige, McCallum, & Winter, 1964). Topographic maps showed that this component was maximal between C3 and Cz. It was measured as the mean amplitude during the 200 ms prior to the shock for the average of these two electrodes. A 2x2 ANOVA of the RP/CNV amplitude revealed a main effect of outcome ($F_{1,14}$=5.00, p=0.042). This arose because preparatory negativity was significantly stronger for go trials compared with no go trials. However, there was no main effect of source ($F_{1,14}$=2.86, p=0.11), and no significant interaction between source and outcome ($F_{1,14}$=0.57, p=0.46).

Evoked responses to the shocks

The average topography of the response to shocks 1 and 2 was examined (see figure 3.5 A). As expected from previous results the average topographical maps show that the response to the both shocks is focused on the central electrodes. Based on this topography, peak amplitudes for the analysis were obtained from the average of electrodes C3, Cz and C4.

Also, the differences between the instructed and intentional conditions for both go and nogo trials were examined (see figure 3.5 B). Topographical maps show large

differences in the left hemisphere, ipsilateral to the shocks but contralateral to the movement. This suggests that this difference is related to motor preparation. The most parsimonious interpretation is that go trials show a stronger (more negative) RP-like movement preparation component for intentional than for instructed trials. This leads to a positivity observed in the electrodes contralateral to movement in the instructed – intentional subtraction.

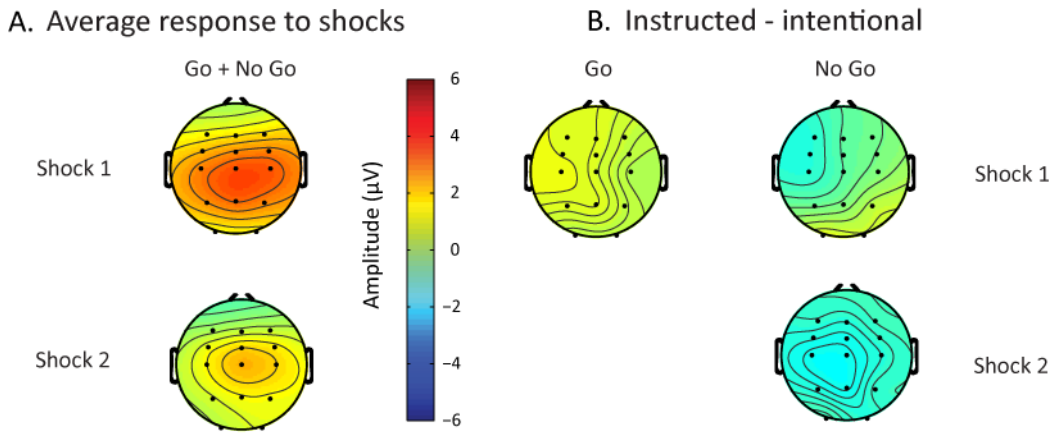


Figure 3.5: Topographical distribution of the neural responses to the shocks. The topographical maps for the response to shock 1 were calculated at 4900 ms, and the response to shock 2 were calculated at 5900 ms from V1 onset. **(A)** Average of all conditions. The topography includes an average of four conditions (instructed *go*, instructed *nogo*, intentional *go*, intentional *nogo*). Shock 2 includes instructed *nogo* and intentional *nogo* conditions only. (**B**) Differences between instructed and intentional conditions, for both *go* and *nogo* trials. Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

Evoked response to shock 1

The topography and peak amplitude of the response to shock 1 is shown in figure 3.6 (highlighted section A in figure 3.4).

A 2 x 2 ANOVA of the average revealed no main effect of outcome ($F_{1,14}$=0.06, p=0.80) nor main effect of source ($F_{1,14}$=0.01, p=0.92) but a significant interaction

effect ($F_{1,14}$=19.433 p=0.001), showing a crossover form in figure 3.6 C. Post-hoc t-tests revealed that the neural response evoked by shock 1 was greater in instructed go than in intentional go trials (t_{14}=3.39, p=0.004). Conversely, the response evoked by shock 1 in instructed no go trials was weaker than that evoked by the intentional no go trials (t_{14}= -2.22, p=0.04).

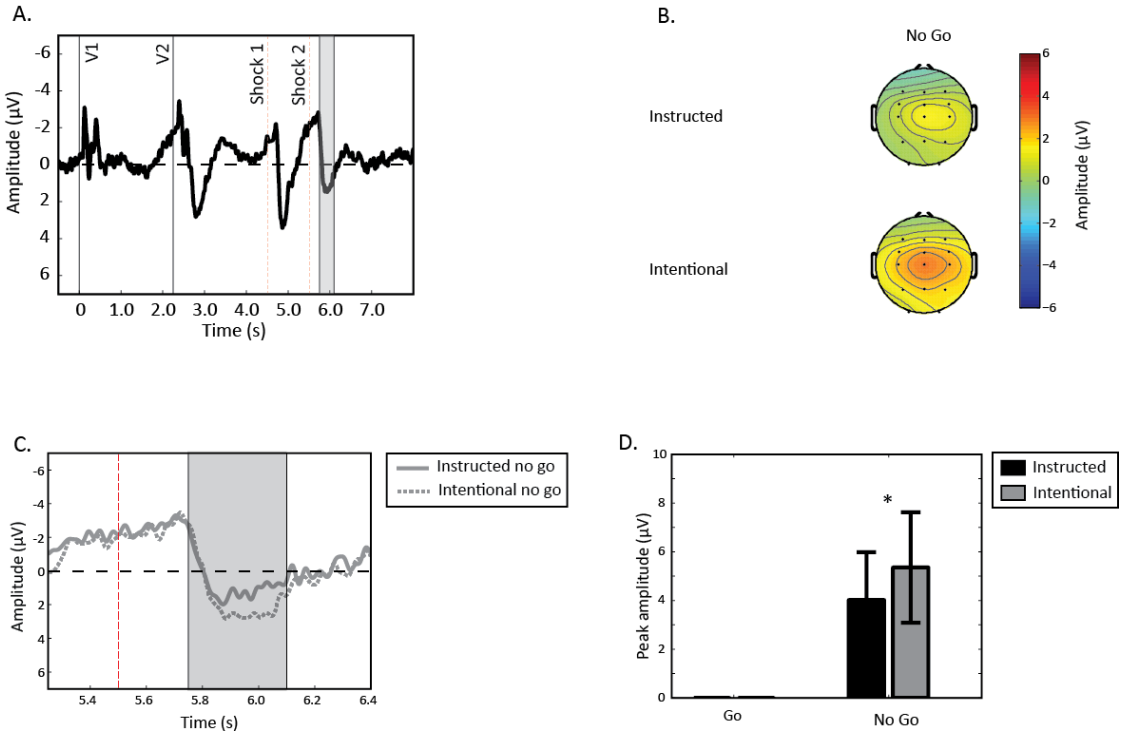


Figure 3.6: Neural response to shock 1 **A.** Time window (4750-5100 ms) in which the amplitude of the response to shock 1 was measured. The vertical dashed line indicates the time of shock onset. **B** Scalp distribution at 4900 ms. **C.** Detail of time window of interest, showing the averaged trace for electrodes C3, Cz and C4 for each condition. **D.** Mean of maximum amplitude for each subject within the selected time window, for the average of electrodes C3, Cz and C4. Error bars show confidence intervals. A significant interaction effect emerged (p<0.01). Post-hoc *t*-tests showed a crossover effect. Shock 1 ERPs were stronger for instructed *go* trials as compared to intentional *go* trials (p<0.05), whereas ERPs to instructed *no go* trials were weaker than those for intentional *no go* trials (p<0.05). Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

Evoked response to shock 2

Because the mean RT to withdraw the left hand was generally shorter than 1 s, shock 2 was generally not delivered in trials where participants made the withdrawal action with their right hand. Hence the analysis of the ERP to shock 2 was confined to no

go trials, in which participants did not withdraw their hand, but resisted the full train of shocks. The topography and peak amplitudes for the no go conditions are shown in figure 3.7 and match that of shock 1 (see also highlighted section B in figure 3.4). A paired t-test comparison revealed a significant difference between the two no go conditions, with the intentional no go again showing a stronger potential in response to itchy shock compared to instructed no go (t_{14}=-2.33, p=0.03).

To compare the neural response to shocks 1 and 2, a 2x2 ANOVA of the no go responses with the factors shock and source of decision was carried out. Results showed no main effect of shock ($F_{1,14}$=1.202, p=0.29), a main effect of decision source ($F_{1,14}$=7.114, p=0.01), with intentional no go trials showing stronger ERP than instructed no go trials. There was no interaction effect ($F_{1,14}$=0.21, p=0.65).

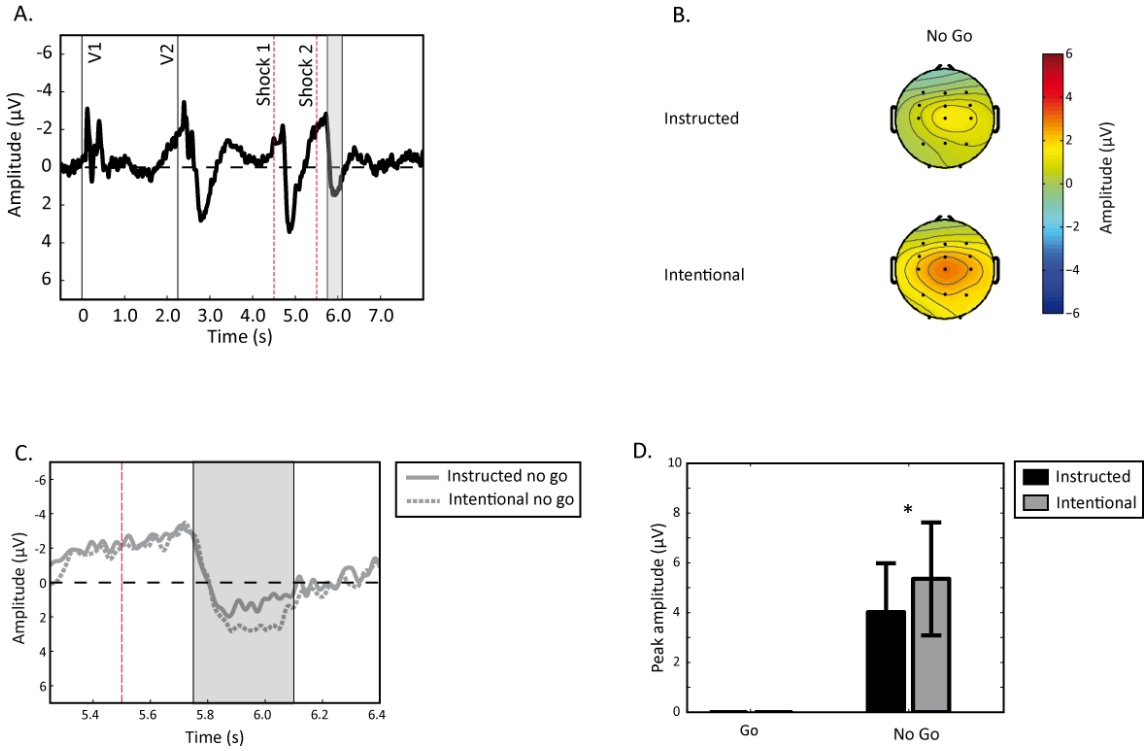


Figure 3.7: Neural response to shock 2. Because shock 2 was only fully delivered in inhibition trials, only *no go* trials are analyzed. **A** Time window (5750-6100 ms) used to measure the amplitude of the response to shock 2. The vertical dashed line indicates the time of shock onset. **B** Scalp distribution at 5900 ms. **C** Detail of time window of interest, showing the averaged trace for electrodes C3, Cz and C4 for each condition. **D** Mean of maximum amplitude for each subject within the selected time window, for the average of electrodes C3, Cz and C4. Error bars show confidence intervals. As in shock 1, instructed *no go* trials evoke a weaker response to shock 2 than an intentional *no go* trials (p<0.05). Adapted from (Filevich & Haggard, 2012), published under Creative Commons Attribution License.

Shock 3 was not analyzed because participants reported at debriefing that they rarely felt it. This reflected a strong habituation, and the evoked potentials were correspondingly weak.

3. 2. 3 Discussion

The findings revealed that the sensory consequences of intentional decisions are processed differently than those of instructed decisions. These differences in processing were found at the neurophysiological level. It is therefore possible that the same differences are observed at the phenomenological level. Experiment 2 was designed to test this hypothesis.

3. 3 Experiment 2

3. 3. 1 Methods

Participants

Twelve participants (7 females, mean age 22 ± 3 years) performed a modified version of the task in Experiment 1. Five participants had taken part in experiment 1, whilst the remaining seven were naïve to the task. Experiment 2 consisted of 5 blocks of 40 trials each.

Task

In a follow-up experiment to the ERP results of experiment 1, participants were required to use a visual analogue scale to rate the unpleasantness of the itchy feeling after each trial (figure 3.8). This had not been done together with the EEG experiment 1 because of time restrictions.

Experimental details were identical to those of experiment 1, apart from two

important differences. First, although the intensity of stimulation was kept constant for the different conditions, participants were told that the stimulation would vary slightly from trial to trial, and they were asked to rate the unpleasantness of the stimulation after each trial. The subjective ratings were reported by a mouse click on a visual analogue scale displayed on the screen.

The duration of the stimulation had a clear effect on how strongly the itch sensation was felt. Therefore, a second difference from experiment 1 was that in order to make the comparison between withdrawal (*go*) conditions possible, the stimulation was always delivered for at least 1s, regardless of whether the participant had withdrawn their hand from the FSR.

Data analysis

The subjective ratings presented large variations over time, regardless of condition. The perceived intensity of the shocks greatly depended on uncontrolled factors, such as habituation, body temperature and alertness. Therefore, the subjective ratings were calculated not in their absolute value but as the deviation from the local average, calculated from the four neighbouring trials.

3. 3. 2 Results

Mean go RTs were very high in experiment 2. Mean (± SD) instructed go RT was 1568 ± 364 ms and mean intentional go RT was 1559 ± 392 ms. These values did not differ significantly (t11=0.19, p=0.855).

Although the same level of stimulation was delivered in every trial, participants were told that each shock was of a slightly different intensity. They were asked to give a subjective rating of the perceived intensity by means of a VAS. Mean ratings for each condition are shown in figure 3.8. A 2x2 ANOVA on the mean reported ratings for each condition revealed a main effect of decision outcome ($F_{1,11}$=6.45, p=0.027), a main effect of source of decision ($F_{1,11}$=5.57, p=0.038) and no significant

interaction effect ($F_{1,11}$=0.009, p=0.928).

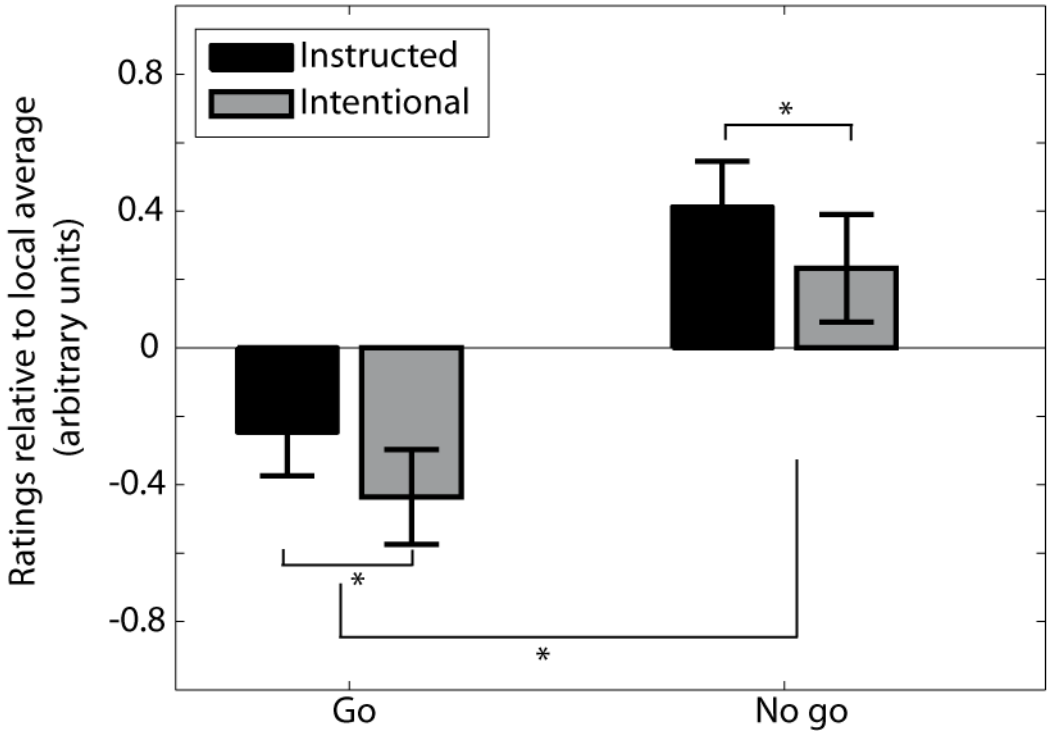


Fig 3.8: Subjective ratings of individual trials relative to the local average of the four neighbouring trials. Error bars show standard error of the mean.

3. 4 Discussion

A paradigm that generates strong urges to make hand withdrawal movements was used to study the inhibitory functions involved in self-control over urges and actions. Further, situations of intentional and instructed inhibition were compared within the same context. Participants were either instructed, or decided for themselves, whether to withdraw their right hand from a response pad, thus terminating a train of unpleasantly itchy shocks delivered via a customised electrode to the left forearm.

This paradigm addressed the consequences of inhibition, following both intentional decisions, and external instructions. Specifically, it was investigated how intentional inhibition influenced neural processing of subsequent itchy shocks, in comparison to

instructed inhibition. In both these conditions, there is no overt behaviour. However the source of the decision to inhibit differs between these conditions. Results show that the source of inhibitory decisions has an important structuring effect on subsequent sensory experience. This study takes the novel approach of addressing the consequences of self-control and intentional inhibition of action on somatosensory processing.

Ecological validity

This experiment used aversive stimulation to induce a strong urge to act. Unlike previous studies of intentional inhibition, it made the choice between action and inhibition motivationally significant. This was combined with the conventional comparison between instructed and intentional decisions for action. The motivational element of the experiment may seem to conflict with the free selection element: if itchy shocks were truly aversive, participants should choose to avoid them. Participants' willingness to accept the experimental instructions perhaps lead them to trade shocks for money (Talmi et al. 2009). However, the balance of positive and negative affect indubitably plays a major role in shaping everyday action choices (Damasio and Dolan 1999). Notwithstanding these general motivations, participants made a fresh choice on every intentional trial, whether to act or inhibit *on that trial*. The random intermingling of intentional and instructed trials was designed to discourage them from preparing sequences of "free" choices extending over several trials. For these reasons, this study may have an ecological validity lacking in previous studies of intentional inhibition.

Experiment 1 - Neural processing of the consequences of intentional decisions

A marked positive-going component occurs in response to each of the three shocks, peaking at around 400ms after shock onset. Previous studies using similar stimuli (Mochizuki et al. 2009) had reported slower positive components, peaking at around 900 ms. Only a speculative explanation can be given for this difference. Mochizuki *et al*. associated the long latency of the evoked response with C-fibre activation,

because of their slow conduction speed. Indeed, itch sensation has been mainly associated with C-fibres (Handwerker, 2010). However, some Aδ-fibres have also been associated with itch in monkeys (Schepers et al., 2008) and humans (Ringkamp et al., 2011). Thus, it can be speculated that the electrical pulses delivered here may also may have stimulated both C-fibres and the faster-conducting Aδ-fibres. Activation of Aδ-fibre can lead to both sharp and burning pain. The joint activation of different fibre populations may have produced the distinct sharp itchy feeling that participants experienced.

In this paradigm, participants always received the first shock, and received the second and third shocks only if they chose to inhibit the action of withdrawing their arm. The third shock was perceived very weakly and accordingly produced only a small ERP. Previous reports (Mochizuki et al. 2009) suggested that diminishing neural activity evoked by itch reflected habituation effects. Decrease in saliency for repeated stimuli has been widely reported (Legrain et al. 2011).

Go and *no go* trials cannot easily be compared directly, because several different factors may contribute to the differences between their neural correlates. Crucially, *go* trials involve motor preparation, whilst *no go* trials do not. No strong conclusions are drawn about differences between *go* and *no go* trials. However, comparing intentional and instructed *sources* of action decisions is possible, within both the *go* and *nogo* condition, because the motor activity is balanced between the two sources of decision.

The neural response for shock 1 produced lower ERP amplitudes when participants intentionally decided to withdraw their hands, as opposed to when they were instructed to do so. The neural response to shock 2 cannot be evaluated in the case of movement conditions, because when the hand has been withdrawn, no further shocks are delivered. However, in the case of no go trials, the neural trace of an aversive stimulus that could have been avoided, but was not, informed about the mechanisms of inhibition. In these trials, the response to the first and second shock can be evaluated. The neural response to both the first and second shocks was significantly *larger* when participants underwent the shocks as a result of their own intentional decision, rather than as a result of an instructed instruction.

Thus, while action trials showed *smaller* ERPs to itchy stimuli in the intentional go trials than in instructed go trials, inhibition trials showed the opposite effect. Stronger ERPs were apparent for intentional no go trials as compared with instructed no go trials. A speculative account suggests that this interaction effect reflects differences in allocation of attention strategies between *go* and *no go* trials. In this paradigm, participants presumably attended to the shocks when they needed to react quickly (*i.e.*, in go conditions). In *no go* conditions, participants may have preferred to "think of something else" and try to completely ignore the shocks. Interestingly, in intentional trials, this effect is reversed. I.e., the attention allocation towards the aversive stimuli in *go* trials, and away from aversive stimuli in *no go* trials would have been less efficient in the intentional conditions, according to this view. It has been suggested (Fleming et al. 2009) that intentional decisions for actions are less definitive, and easier to change, than instructed actions. Could this explain the differences found between the processing of aversive stimuli in intentional and instructed no go trials? If Fleming *et al* are correct that intentional decisions to act or inhibit still leave open the counterfactual possibility, it may be speculated that unpleasant consequences of intentional decisions might be strongly processed because of feelings of regret for the missed opportunity of doing otherwise.

In other words, instructed trials have clear instruction, and there is a clear correct response. This is not the case with intentional trials, in which any course of action would be correct. Therefore, in line with results reported by Fleming et al (2009) attention allocation in intentional trials may represent an intermediate situation between the two extremes: allocation of attention towards an aversive stimulus in instructed go trials, and allocation of attention away from an aversive stimulus in instructed no go trials.

An alternative, but closely related interpretation, relates to motor processing. In this study we aimed at investigating action inhibition *indirectly* by addressing the sensory processing of its consequences. However, motor and sensory processes were not temporally segregated in our task. Moreover, EEG techniques do not allow us to unequivocally identify the sources of the modulation of the shock components as either clearly sensory, or clearly motor. Thus it remains possible that our results reflect movement-related processing. Because we did not investigate the periods

before action directly, or EEG components that are classically related to action, the hypothesis that our peak ERP amplitude is affected by motor processes remains speculative. The influence of motor preparation on the shock component amplitude is unclear. Importantly however, this interpretation remains compatible with the "intermediate" account of intentional decisions suggested above.

Intermediate ERP peak amplitudes may therefore reflect intermediate levels of motor preparation. In particular, whereas instructed *go* trials are associated with a high levels of motor preparation, intentional *go* trials seem to present lower levels of motor preparation, closer to the *no go* conditions. In turn, instructed *no go* trials are presumably associated with lower levels of motor preparation because no action should occur in *no go* trials. Intentional *no go* trials present higher levels of action preparation, closer to *go* trials.

Crucially, in line with the results reported by Fleming et al (2009), intentional decisions for action may represent situations that are *less* committed to than instructed decisions, and therefore may be easier to change than corresponding instructed decisions. If the attention allocation account is correct, our results may interestingly extend this interpretation from the purely motor processing addressed by Fleming et al. to the sensory processing of decision consequences.

Experiment 2 - subjective experience of the consequences of intentional decisions

Motivated by the ERP results, experiment 2 aimed at providing the phenomenological counterpart of the findings of experiment 1.

The mean subjective ratings were compared fro the four experimental conditions. Statistical analyses revealed a main effect of decision outcome. This effect is trivial, and related to the duration of the stimulation received. Go (withdrawal) trials involved in average around 1s of electrical stimulation, whereas no go trials involved at least 2s of perceivable stimulation. Because the perceived intensity accumulated over time, no go trials led to stronger subjective feelings of the itch stimulation.

Based on the ERP findings from experiment 1, intentional inhibition trials were

expected to elicit a higher subjective report of itch, whilst intentional action trials would be associated with lower ratings of itch. This interaction effect observed in the ERP amplitude was not mirrored by the reported subjective experience. Instead, participants rated itchy stimulation to be lower in both intentional conditions, and not only in the intentional action condition.

The results of the two experiments are not easy to reconcile. Taken together, the results show that the neural response evoked by shocks 1 and 2 are not the only factors contributing to the subjective experience of itch in this paradigm, as reported on a VAS. Other possible factors influencing itch perception are folk knowledge about the possibility of choosing (present only in the intentional conditions). It is a well-known phenomenon that human participants tend to rate favourable outcomes as more favourable when they arise as a consequence of their own choice (Mellers, 2000). In the same way, participants tend to judge aversive stimuli as less aversive when they have some control over their intensity and timing (Staub, Tursky, & Schwartz, 1971). Participants may have been biased to believe that freedom of choice should lead to decreased aversive experiences. In addition, it could be speculated that the ratings are very post-perceptual and attributional. They are outputs from the inferential narrative brain. The ERPs on the other hand reflect processes of the decisional, sensory brain.

3. 5 Conclusion

Inhibition of action can take two rather different forms depending on its time-course. First, it can be a rather tonic behavioural control. This form corresponds to the everyday concept of willpower (Baumeister, Schmeichel, & Vohs, 2007; Vohs & Schmeichel, 2003). For example, someone who exerts self-control over their eating behaviour may need to continuously inhibit the urge to eat. These inhibitory processes are continuous and ongoing, rather than discrete and precisely-timed. Inhibitory self-control may also appear in a more phasic form, as a last-minute

inhibition of specific and discrete action impulses (i.e., veto) (Libet, 1999). The type of inhibition required in this task lies somewhere on the continuum between these two extremes. Because some trials required inhibition and others required a quick action, the task was designed to encourage phasic, discrete inhibition, rather than generalized, tonic self-control. The present results therefore address the concept of self-control or "willpower" in a novel experimental way, and suggest that self-control may have different neural bases than instructed inhibition.

Chapter 4 Antecedent brain activity predicts intentional *whether* decisions

Decisions to inhibit are alleged to have a unique relation to conscious thought and cognitive control. This EEG experiment examined the unconscious precursors of an intentional decision to inhibit. The results revealed that prior neural activity could bias intentional decisions to act or transiently inhibit action. "Free" decisions to inhibit action may be unconsciously caused. Like other cognitive control processes, intentional inhibition takes place against a backdrop of ongoing neural activity

4. 1 Introduction

Previous studies have linked preparatory activity preceding voluntary action to decisions about *what* action to make e.g., (Deiber et al., 1991; C. D. Frith, Friston, Liddle, & Frackowiak, 1991; Jueptner, Frith, Brooks, Frackowiak, & Passingham, 1997), or *when* to make it (Jahanshahi et al., 1995; Libet, Wright Jr., & Gleason, 1982). Both these components were shown to have unconscious neural precursors. The readiness potential (RP) is an accepted marker of neural preparation for action (Dirnberger et al., 1998; Jahanshahi et al., 1995). Libet (Libet et al., 1983) famously identified RPs already occurring around 200 ms prior to the conscious decision to move (*when* component). Soon et al., (Soon et al., 2008) found that brain activity several seconds before conscious decision could predict which hand people chose to act with. However, the decision about *whether* to act has received less attention. Such decisions can be taken at almost any stage during motor preparation, up until a point of no return (Logan et al., 1984). Libet controversially suggested that last-minute decisions to inhibit action may involve a purely conscious form of "free won't". But theoretical grounds suggest that conscious decisions to inhibit must depend on unconscious brain processes, just like decisions to act (Velmans, 2002). However, neural precursors of voluntary inhibition have not yet been identified experimentally.

In this study, participants had either to make a rapid key press action, or *transiently* inhibit executing the key press, so as to briefly delay their response. In this way, action inhibition was operationalized as a transient process, characterized by delayed responding, rather than as a complete suppression of all behavioural output. This operational definition has the advantage of matching the action and inhibition conditions more closely, since both conditions include a motor response – though with differing latencies. In everyday life, such impulse control by delaying an intentional response may help in accumulating further information about the environment prior to responding (Shadlen & Newsome, 1996), or in synchronising a joint action (Sebanz, Bekkering, & Knoblich, 2006).

The neural activity preceding intentional decisions to act was compared to the neural activity preceding intentional decisions to briefly inhibit action. In intentional conditions, participants were not explicitly instructed whether to act rapidly or to

delay in any given trial, but rather chose for themselves. The hypothesis was that, in the absence of any external instruction to act rapidly or inhibit, some other factors, such as transient fluctuations in participants' brain states, may be relevant to their decision. Therefore, the intentional conditions would provide a situation in which putative internal fluctuations could lead to an overt modification of behaviour. The rationale was also that external instructions about action would produce a stronger drive of neural activity, overriding any intrinsic fluctuations. Consequently, the levels of neural activity preceding external instructions to act or briefly inhibit action were compared. Several recent studies suggest that the instantaneous state of the brain at the time when a new information-processing operation begins can play an important role in how information is processed. For example, the probability of remembering an item depends on preceding electrical neural activity (Otten, Quayle, Akram, Ditewig, & Rugg, 2006), and the probability of detecting a visual stimulus depends on the phase of EEG alpha rhythm over posterior brain regions (Busch, Dubois, & VanRullen, 2009). By analogy, it was hypothesized that intentional decisions to act or inhibit would depend on the progression of preceding activity in the brain.

In this way, this experiment followed a factorial design in which the *differences* in neural activity between intentional decisions to act and intentional decisions to inhibit were compared with the *differences* in neural activity between instructed decisions to act and instructed decisions to inhibit. It was therefore assumed that sorting trials according to action or inhibition could reveal patterns of preceding neural activity that might putatively bias the outcome of intentional decisions. In instructed decisions, in contrast, the cause of the decision to act or inhibit is assumed to lie in the imperative stimulus, rather than any putative pattern of preceding neural activity.

EEG activity was therefore measured around the time of an external instruction to either act quickly or delay transiently an action, or around the time of a cue that invited participants to intentional decide to either act quickly or delay transiently. Although ERP methods do not typically provide high spatial resolution, they do provide high temporal resolution, (Luck, 2005). This makes ERP methods particularly suited for these purposes, as they allowed for an identification of the

neural activity preceding an instruction that influenced intentional decisions in response to the instruction.

4. 2 Methods

Fourteen naïve healthy volunteers (9 females, mean 24 years, 12 right-handed) participated in this experiment. Procedures were approved by the UCL research ethics committee and were in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants involved in the study.

Before further data analysis, one participant was excluded from EEG analyses due to excessive eye blinking, leaving a total of 13 participants. Each participant performed 8 blocks of 70 trials each, yielding a total of 560 trials.

Each trial belonged to one of five possible experimental conditions. Trials could belong to one of five possible trial types. 4 or these trials types were part of a factorial design, namely the *instructed rapid*, *instructed delayed*, *intentional rapid* or *intentional delay* trials. An additional *nogo* condition was included.

Each trial began with a variable fixation cross period (500 to 1200 ms, see fig 4.1). A warning sign (a grey circle subtending 1.5º, duration 200 ms) appeared first. The fixation cross reappeared for 500 ms and was followed by an instruction cue (a coloured circle, 1.5º visual angle, 200 ms duration). The instruction cue indicated one of four main possible trial types. In the instructed rapid condition (240 trials, 43%), participants were asked to press a key with their index finger as quickly as possible. In the instructed delayed condition (80 trials, 14%), participants had to make the same movement but with the "shortest possible delay". The exact duration of the delay was not explicitly specified to the participants, but they were encouraged to delay their action for a period of time that was "as short as possible". In the intentional conditions (160 trials, 28%), participants saw a cue that indicated that they were free to choose which action outcome to take. Namely, immediately upon the appearance of the intentional cue, participants were asked to decide freely whether to act rapidly or after the shortest possible delay. In this way, the experiment followed a 2x2 factorial design, with the factors source of decision

(instructed/intentional) and outcome (rapid/delayed). The percentages of trials were constant across all blocks.

An additional *nogo* condition was included, in which participants were asked to refrain from acting (80 trials, 14%). This condition was intended to make the task more demanding and prevent drifts of attention. The neural signals associated with these trials were not analyzed.

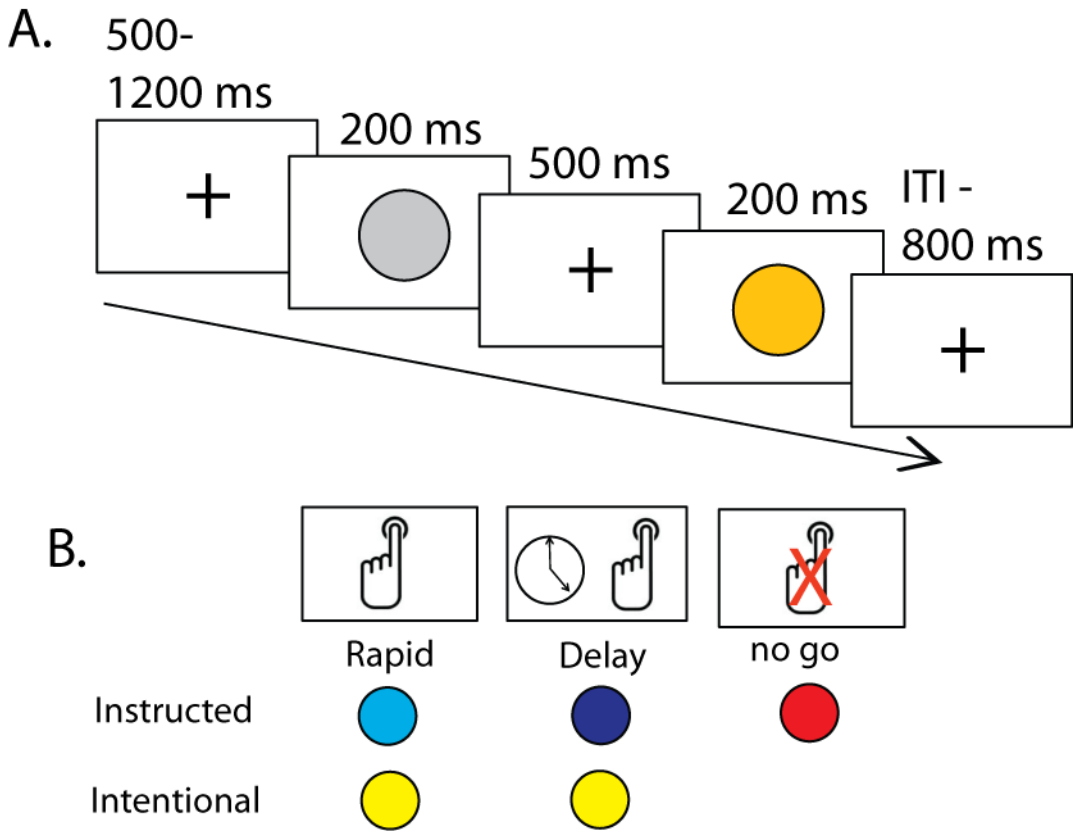


Figure 4.1: A. General timing of the task A variable fixation cross period (500 – 1200 ms) was followed by a brief (200 ms) presentation of a warning sign. The fixation cross reappeared and 500 ms after the offset of the warning sign, an instruction cue appeared on the screen (200 ms). The instruction cued participants to press a key either rapidly, or with a short delay, or to intentionally decide between rapid and delayed pressing. **B.** The task followed a factorial design. The instruction cued participants to press a key either rapidly, or with a short delay, to intentionally decide between rapid and delayed pressing. In an additional nogo condition participants were instructed to refrain from acting at all. This *nogo* condition was included to discourage attentional drifts. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

The rationale behind the experimental design was as follows. Neural networks continually exhibit small fluctuations in state, which may have significant effects on behaviour (Fox, Snyder, Vincent, & Raichle, 2007). These effects may be particularly relevant for behaviour in the absence of other clear, strong external signals. Here, the aim was to identify possible effects of such intrinsic fluctuations on the intentional choice between action and transient inhibition. It was assumed that similar intrinsic fluctuations should exist before instructed choices to act rapidly or delay actions. However, the strong signals linked to external instruction should override these weak internal signals, so that no differences between activity preceding actions and inhibition should remain. Therefore, in a factorial design the neural activity prior to decisions to act rapidly was compared to the neural activity of decisions to delay action, where this decision was either intentional or instructed. Differences in the neural activity preceding intentional action decisions were expected, because the preceding neural activity should strongly influence the intentional decision between different action outcomes. Further, no differences were expected in the preceding activity between rapid and delayed instructed action, since the instructional signal should then have a far stronger influence on behavioural outcome.

In experiments involving intentional inhibition, there is a high risk of participants deciding in advance not to make an action (Brass & Haggard, 2007). In cases of early decisions not to act, no action will be prepared, and consequently no action inhibition will be necessary. Therefore, tasks addressing intentional inhibition should encourage action preparation. In this case, a high number of rapid instructed trials were included to encourage action preparation, to make delaying effortful, and to discourage participants from deciding in advance whether to respond rapidly or inhibit and delay on intentional trials. Further, only for trials in the instructed rapid condition, participants were rewarded (3p) for every key press that was faster than their average in the previous block. The experimental design was therefore not strictly balanced, but emphasized the need for true action inhibition.

In intentional trials, participants were asked to balance their choices between rapid and delayed responses. The hand used for responding was fixed for each block, and

alternated between blocks. The correspondence of colours to instructions rotated across participants, and was additionally reversed for each participant for the second half of the experiment. Trials within each participant were randomized, but the proportion of trial types was valid for each block.

Data analysis

Trials with RTs below 200 ms were rejected, as potentially anticipatory (1.76 ± 1.67%). The average commission error rate in *nogo* trials was 10±0.9%. These trials were included merely to engage attention, and to ensure that participants responded only after receiving the go signal. *Nogo* trials were not further analyzed. RTs for each condition were analyzed in a 2x2 repeated measures ANOVA, with the factors decision source (instructed/intentional) and response speed (rapid/delay).

To examine whether participants followed any obvious strategy to produce a balanced outcome between rapid and delayed intentional trials, the run length was evaluated in the response sequences in intentional trials. This experiment consisted of 8 blocks, with 20 intentional trials each. For each block, the instructed trials were excluded and the length of runs was measured (i.e., sequences of uninterrupted repetitions of the same outcome) for each participant. If participants had been producing obvious sequences such as 'AABBAABB' they would produce a single run-length only (in this case, a run-length of 2). 8 independent "blocks" of 20 "trials" each were then randomly generated by sampling without replacement (using the *randsample* function for Matlab, the Mathworks, Inc) from a population of 10 quick trials and 10 delay trials. A balanced number of simulated quick and delayed trials was necessary because in the intentional rapid and delayed conditions were identified on the basis of a median split.

EEG data acquisition and analysis

A SynAmps amplifier system and Scan 4.3 software (Neuroscan, El Paso, TX) were used to record EEG data. Activity from fourteen scalp electrodes was recorded (F3,

Fz, F4, FC3, FCz, FC4, C3, Cz, C4, P3, Pz, P4, O1, O2) and the right and left mastoids. The scalp electrodes were placed according to the international 10-20 system. The reference electrode was AFz and the ground electrode was placed on the chin. All electrode impedances were kept below 5 KΩ. Electroculograms (EOG) were recorded from bipolar electrodes placed on the left and right external canthi (to detect horizontal eye movements), and on the right supra-orbital and infra-orbital positions (to detect vertical eye movements). EEG signals were amplified and digitized at 500 Hz.

EEG data were analyzed with EEGLAB software (Delorme and Makeig, 2004). Data were first re-referenced to the linked mastoids. Data were digitally band-pass filtered between 0.05 Hz and 30 Hz. Continuous EEG data was time-locked to the instruction stimulus, and epochs were defined from -850 ms to 700 ms after the instruction sign. A baseline period was defined for each epoch from -850 to -700 ms (between 0 and 150 ms prior to the onset of the warning signal). The hand required for action was alternated and specified at the beginning of each block. Lateral (non-midline) electrodes were inverted in the right hand blocks, as if all data had been collected from the *left* hand. Because the lateral electrodes from the right hand blocks were inverted, electrodes in the left hemisphere are now ipsilateral to action. Similarly, electrodes in the right hemisphere are contralateral to action.

Right-left hand symmetry cannot be assumed in this situation. First, the left hemisphere is dominant for action preparation (Bradshaw, 2001). Second, whereas RPs associated with right hand movements are normally distributed, this is not the case for left hand movements (Dirnberger, Duregger, Lindinger, & Lang, 2011). The distribution of the early left-hand movement RP amplitudes shows negative skewness values, even in cases of very simple actions, such as key presses with the index finger. This suggests that movements with the non dominant hand may require more attentional resources and/or special preparatory processes.

To remove blink artefacts, epochs were rejected if the difference between the two vertical EOG channels was larger than 90μV.

For ERP data analysis, three consecutive 50 ms time windows prior to the instruction cue were defined (-150 to -100 ms, -100 to -50 ms and -50 to 0 ms). These timepoints were selected based on previous studies on prestimulus ERP activity

(Otten et al., 2006). The mean EEG amplitude in the electrode Cz was calculated for each participant. As in the case of the RTs, mean window ERP amplitudes were analyzed in repeated measures ANOVA. Greenhouse-Geiser (GG) corrections were applied when appropriate, but full degrees of freedom are reported.

The experimental design included a much larger number of trials in the instructed rapid condition than in the other three conditions. Consequently, resampling methods were applied to control for uneven numbers of trials (Gruber & Otten, 2010). For each participant, the number of trials in the instructed delay condition was found. The same number of trials was then randomly sampled, with replacement, from the trials in the instructed rapid condition. These two populations of trials were then combined to get an overall distribution of instructed RTs. A trial was considered as "correct" in the instructed rapid condition if its RT was quicker than the median of the distribution of instructed RTs. In the same way, a trial was considered as correct in the instructed delayed condition if its RT was slower than the median of the distribution of instructed RTs. Finally, the mean CNV amplitude measured from electrode Cz was obtained for all four main trial types, in each of the 50 ms time windows prior to the instruction, and averaged across subjects. This procedure was repeated 10,000 times.

4. 3 Results

Behavioural results

Following the monetary reward incentive to the instructed rapid trials, participants became quicker in each block. The total number of rewarded trials (i.e., those instructed rapid trials that were quicker than the average of the previous block) was 156 ± 10 (mean ± SD), and there was a mean decrease in RT of 55 ms.

Instructed trials were classified as rapid or delayed *a priori*, according to the instruction given in each trial. Intentional trials lacked a specific instruction, and hence were classified as rapid or delayed *a posteriori*, on the basis of a median split of each participant's intentional response RT distribution (see appendix A for individual distributions, and see below for sensitivity analysis). Because the

intentional trials were classified as rapid or delayed on the basis of a median split, exactly half of the trials were rapid, and half of the trials were delayed.

To determine the effect of the decision in the intentional conditions, a 2x2 ANOVA was performed on the RTs with the factors decision source (instructed/intentional) and response outcome (rapid/delay). The mean RTs are shown in figure 4.2. The main effect of source of decision ($F_{1,13}$=7.15, p=0.019) arose because intentional responses were slower than instructed responses. This suggests that intentional decisions to respond rapidly or to transiently inhibit and delay involved a time-consuming cognitive process occurring after the cue. The main effect of outcome ($F_{1,13}$=81.43, p<0.001) unsurprisingly showed that participants significantly delayed their RTs both in instructed and in intentional conditions. The interaction between source of decision and outcome was not significant ($F_{1,13}$=0.12, p=0.734).

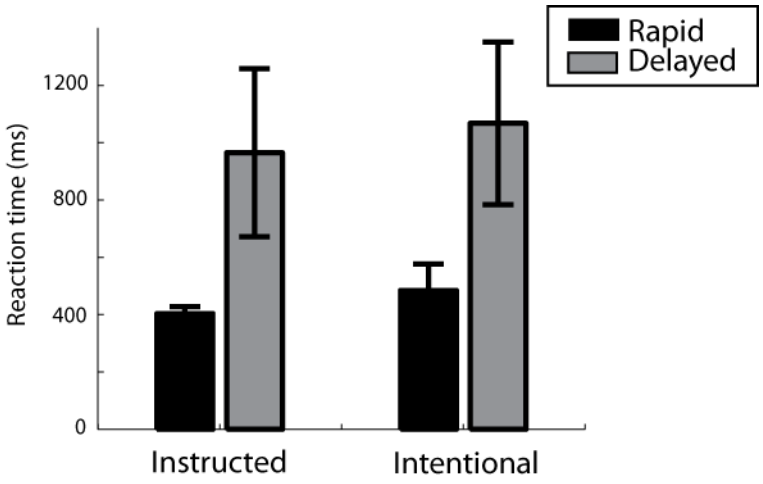


Figure 4.2: Delay trials have longer RTs than rapid trials, as expected by the task instructions. More importantly, intentional trials are slower than instructed trials, suggesting a intentional decision process occurring after the visual cue. Instructed rapid and delayed trials were classified on the basis of the instruction provided. Intentional rapid and delay trials were classified on the basis of a median split (see text for details). The two methods produce a similar RT separation. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

Participants switched hands in each block. Therefore there could have been an effect of hand used on the mean RTs. To examine this possibility RTs for blocks in which

participants used their dominant *vs.* nondominant hand were compared. A repeated measures three-way ANOVA was conducted, with the factors block subset (dominant hand/nondominant hand blocks), source (instructed/ intentional) and outcome (rapid/delayed). The results show a main effect of block subset ($F_{1,12}$=6.40, p=0.026), indicating that participants were quicker to make actions with their dominant hands, as might be expected. However, there was no three-way interaction ($F_{1,12}$=0.08, p=0.770), indicating that the hand used did not affect the interaction of source x outcome that is of interest here.

Similarly, to rule out low-level effect of the physical stimuli, the correspondence between the visual cue colour and the instruction was changed half way through the experiment. This could have led to a significant Stoop-like effect (Stroop, 1935) that affected mean RTs. To explore this possibility, a repeated measures three-way ANOVA was conducted, this time, with the factors block subset (first half/second half), source (instructed/ intentional) and outcome (rapid/delayed). Results show a main effect of block subset ($F_{1,12}$=18.38, p=0.01), suggesting that participants had learnt the association in the first half of the experiment, and the switch in association between colour of the instruction cue and the task generated a Stroop-like effect. Importantly however, the three-way interaction was not significant in the case of hand used ($F_{1,12}$=0.47, p=0.505), suggesting that the source x outcome interaction was not modulated by Stroop-like effects.

Participants' strategies

Participants were asked to produce roughly 50% rapid and 50% delay responses. This may have led to stereotyped behaviour, such as chunking or direct alternating strategies. If this had been the case, the decision to act rapidly or delay would not have been taken just before the instruction, but presumably at the onset of the trial. To discourage this strategy, instructed and intentional trials were interleaved. The alternation between intentional rapid and delayed trials would therefore require a higher effort of maintenance of the preceding history of choices in working memory. To formally test that this was not the case, possible specific simple chunking strategies (e.g., AABBAABB) were sought, with the standard method of run-length analysis (Nickerson & Butler, 2009; Wald & Wolfowitz, 1940).

The run length in each participants' sequence of intentional responses was examined (after discarding the interleaved instructed responses), and compared with simulated random data (see figure 4.3).

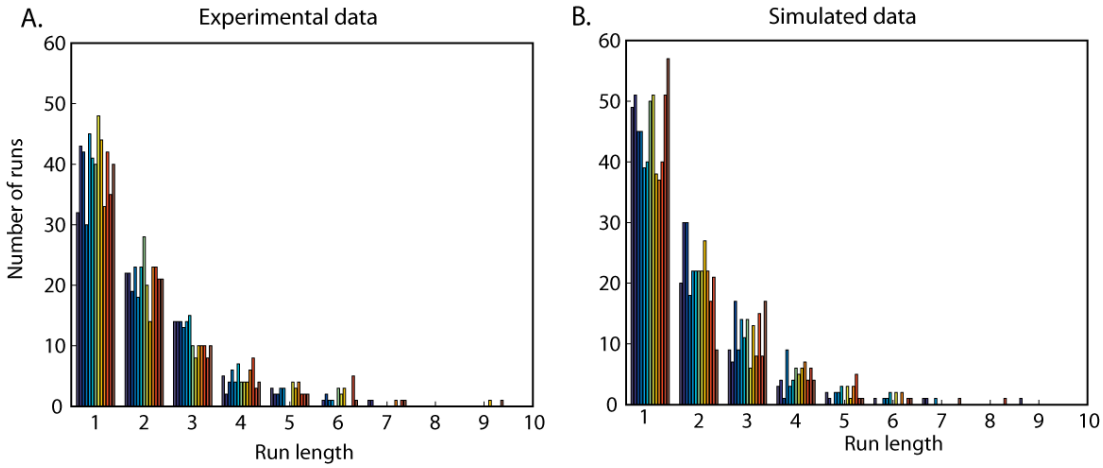


Fig 4.3: Distribution of run lengths in the intentional trials for **A** all participants and **B** simulated random data. A run is an uninterrupted sequence of repeated choices to either press rapidly or delay. Each colour represents an individual participant. Participants tended to produce less runs of length 1 (less direct alternations) than would be expected from random sequences. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

The simulated data shows the same pattern than the experimental data. To test if this was indeed the case, a 2x4 ANOVA was done with the factors data type (experimental/simulation) and run length (1 to 4). The results showed show a main effect of data type ($F_{1,13}$ =6.98, p=0.02) and a significant data type x run length interaction ($F_{2,26}$=4.19, p=0.019). Consequently independent paired t-tests for the number of runs of length 1,2,3 and 4 were conducted. The results showed that only the number of run lengths of 1 differed significantly between the experimental and the simulated data (experimental data mean ± SD: 40 ± 6 runs; simulated data: 46 ± 6 runs; t_{13}=-2.981, p=0.010). Participants showed fewer runs of length 1 than expected based on simulation results, indicating that subjects tended to avoid direct alternation (ABAB).

Finally, to statistically test for randomness all blocks of each participant were collapsed into a single run of 160 trials, a Wald-Wolfowitz Runs test was performed

(Wald & Wolfowitz, 1940) for each subject. The null hypothesis that the sequence generated was random was not rejected for any participant (all $p>0.135$).

ERP results

After artefact rejection, an average of 164 ± 70 trials (69% of original trials) remained for the instructed rapid condition, 67 ± 12 (84%) trials for the instructed delayed condition, 57 ± 22 (72%) trials for the intentional rapid condition and 69 ± 9 (86%) trials for the intentional delayed condition. Participants were told that they could blink only after having made an action, to prevent the common tendency to blink and press the key at the same time. This instruction might have potentially introduced the observed in the percentage of rejected trials between action and transient inhibition conditions. Because key presses occurred earlier in the rapid conditions, this could have been the reason for higher proportion of more blinks occurring during the epoch of interest.

Event-related potentials (ERPs) showed a clear negativity before the instruction signal (figure 4.4). This corresponds to the classical contingent negative variation (CNV), (Tecce, 1972; Walter et al., 1964).

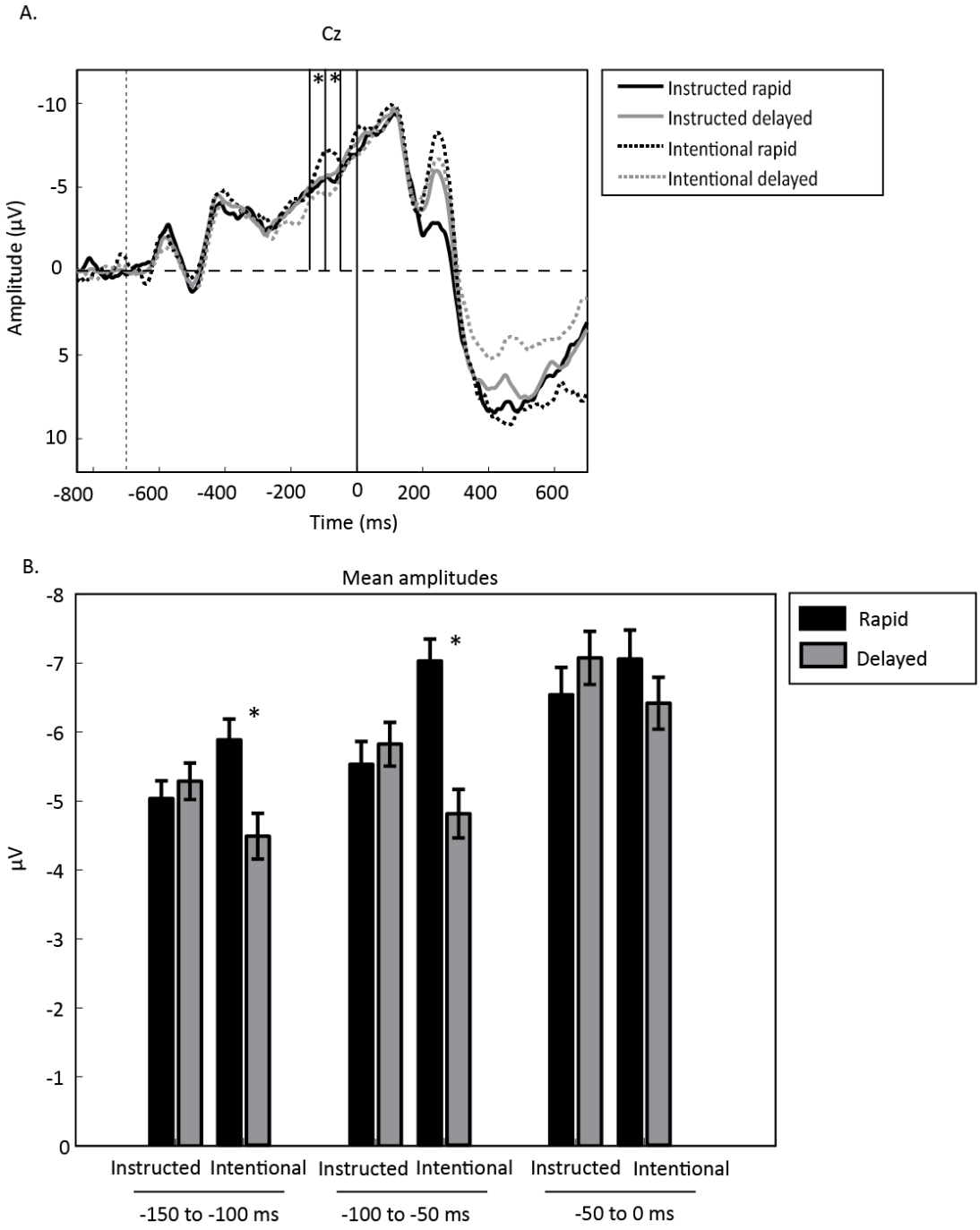


Figure 4.4: A. Averaged CNV amplitude in electrode Cz for the four main conditions, time locked to the appearance of the instruction cue (time 0 corresponds to the onset of the instruction cue). Note the difference in CNV amplitude between two intentional rapid and delayed trials (solid lines), but no difference in CNV amplitude between instructed choice rapid and delayed trials (dashed lines). Asterisks indicate a significant ANOVA interaction (*F* test, $p<0.05$, uncorrected). Vertical dashed line at -700 ms indicates onset of warning signal and the end of baseline period (-850 to -700 ms). **B.** mean CNV amplitudes for the three time windows considered, in electrode Cz. Asterisks in panel B indicate a significant difference in the mean amplitude in the time window (t-test, $p < 0.05$, two tails, uncorrected). Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

To examine the topography of this component, scalp maps were produced in the three time windows of interest. These maps show that the CNV shows a broad distribution, centred on electrode Cz (see figure 4.5).

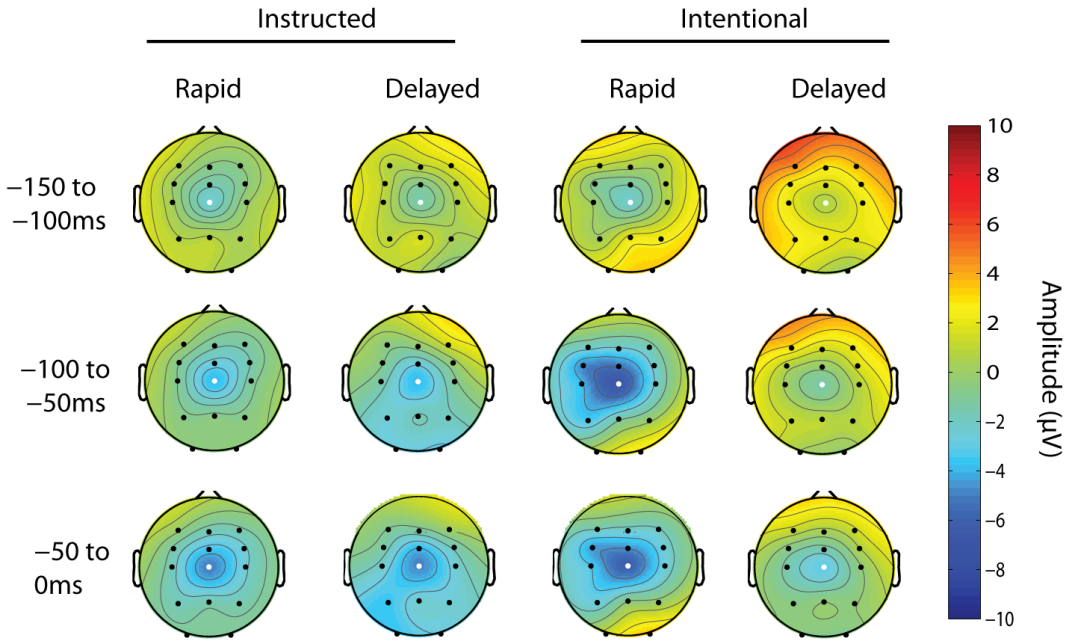


Figure 4.5: Topographical distribution of the CNV component for each of the four main conditions, averaged over three time windows selected for analysis. White highlight shows electrode Cz, from which the mean time window amplitudes were obtained for statistical analyses. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

To explore differences in the CNV amplitude between conditions, the topography of the CNV potential was first explored. A 2x2x3 ANOVA was conducted, with the factors source (instructed/intentional), outcome (rapid/delay) and electrode group (ipsilateral/midline/contralateral). Segmenting electrodes into regions rather than entering them individually as factors into an ANOVA is a more informative approach (Luck, 2005). The parietal and occipital electrodes were excluded, given the a priori hypothesis of the known topographical distribution of the CNV (Tecce, 1972; Walter et al., 1964). To simplify the analyses, the single time window of -150 to 0 ms prior to the instruction cue was considered.

Results show a main effect of electrode group ($F_{2,24}$=8.59, p=0.002), no main effect of source ($F_{1,12}$=0.03, p=0.874) and no main effect of outcome ($F_{1,12}$=0.95, p=0.348). There was a marginally significant source x outcome interaction effect ($F_{1,12}$ =4.55, p=0.054). This effect did not interact with electrode group ($F_{2,24}$=0.24, p=0.673). Therefore the standard approach was taken of using the electrode Cz for the analysis of the CNV amplitude.

Next the possibility that there were any differences between conditions over the three time windows defined for analysis was explored. A 2x2x3 ANOVA with the factors source, outcome and time bin (-150 to -100 ms/-100 to -50 ms/-50 to 0 ms). Results show a main effect of time window ($F_{2,24}$=8.77, p=0.007); no main effect of source ($F_{1,12}$=0.03, p=0.862) and no main effect of outcome ($F_{1,12}$= 1.16, p=0.302). There was a significant interaction effect between source and outcome ($F_{1,12}$=6.06, p=0.030). This interaction was explored by post-hoc testing. In the intentional condition, the CNV amplitude measured from Cz was reduced (i.e., less negative) when participants chose to transiently inhibit and delay than when they chose to respond rapidly. In contrast, the instructed condition showed no difference between rapid and delay trials. That is, CNV amplitude just before the decision cue had a specific association with subsequent intentional decisions to respond rapidly or to delay. Figure 4.6 shows the topographical distribution of these differences.

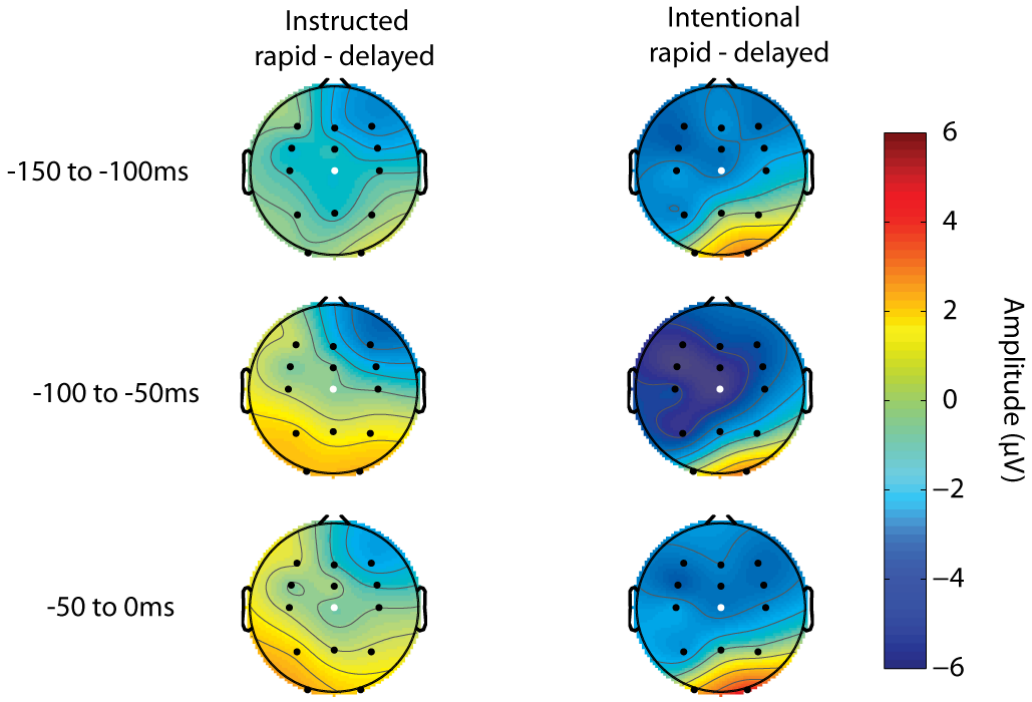


Figure 4.6 Topographical distribution of the difference in brain activity between rapid and delayed trials. Depicted values are averaged amplitudes over 50 ms time windows. There are stronger differences in intentional than in instructed conditions. White highlight shows electrode Cz, from which the mean time window amplitudes were obtained for statistical analyses. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

This two-way interaction in turn shows a marginally significant interaction with time window, as shown by the three-way interaction effect in the ANOVA analysis, $F_{2,24}$=3.8, p=0.051). Because of this marginal three-way interaction, the source x outcome interaction effect was evaluated in each one of the three time windows (-150 to -100 ms, -100 to -50 ms and -50 to 0 ms). Results show a source x outcome interaction in the -150 to -100 ms (p=0.041) and -100 to -50 ms window (p=0.016), but not the -50 to 0 ms window (p=0.110), see tables 4.1 and 4.2. Because post-hoc t-tests were examined only to explore significant interactions, corrections for multiple comparisons were not used.

Table 4.1: Mean (± SD) ERP amplitudes in the three prestimulus intervals considered.

	Amplitude (μV) ± SD			
	Instructed		**Intentional**	
Time interval (ms)	**Rapid**	**Delayed**	**Rapid**	**Delayed**
-150 to -100	-5.03 ± 3.35	-5.28 ± 3.42	-5.87 ± 3.93	-4.47 ± 4.30
-100 to -50	-5.53 ± 4.33	-5.82 ± 4.11	-7.01 ± 4.14	-4.80 ± 4.58
-50 to 0	-6.54 ± 5.16	-7.07 ± 5.03	-7.04 ± 5.47	-6.40 ± 4.89

Table 4.2: Results of statistical analyses of EEG amplitudes in three 50 ms time bins. Interaction term of a 2X2 ANOVA (source of decision x outcome), and results of the follow-up t-tests. All p values are uncorrected. See text for details.

	Interaction		**Intentional rapid – delayed**		**Instructed rapid – delayed**	
Time interval (ms)	$F_{1,12}$	**p**	t_{12}	**p**	t_{12}	**p**
-150 to -50	5.18	0.041	-1.79	0.097	0.58	0.56
-100 to -50	7.75	0.016	-2.57	0.024	0.66	0.515
-50 to 0	2.96	0.110	-0.80	0.437	1.05	0.313

Rapid and delayed trials were classified *a priori* in the instructed conditions, but *a posteriori* in the intentional conditions. It was thus assumed that *instructed* rapid and *instructed* delayed trials were drawn from separate populations, with different mean RTs. However, if participants had completely failed to follow the instruction to respond rapidly, or with a delay, then instructed rapid and instructed delay RTs

would not have differed. In the CNV, RT and ERP amplitude have been shown to be inversely related (Hillyard, 1969). Therefore, this could also have suppressed differences between instructed rapid and instructed delayed ERP amplitudes. The interaction found between instructed and intentional conditions could then be an artifact of using *a priori* classification criteria for instructed conditions, but *a posteriori* classification criteria for intentional conditions.

The present results suggest that this is not the case, for several reasons. First, a strong main effect of outcome emerged when instructed trials were classified *a priori* according to the instruction signal, suggesting that participants indeed attended to the instruction to respond rapidly or to delay, and indeed generated two distinct populations of instructed trials with minimal overlap in RT. Crucially, there was no significant interaction ($p=0.73$) between decision outcome and decision source. This was also the case when controlling for hand used or possible confusion due to Stroop-like effects. These findings suggest that participants were equally able to produce distinct rapid and delayed actions in instructed and intentional conditions. Thus, the *a priori* criterion for instructed trials and *a posteriori* criterion for intentional trials were approximately matched. Since treatment of RTs was successfully matched across instructed and intentional conditions, differences between ERP amplitudes cannot simply be a consequence of differences in RT distributions.

Second, an additional analysis was performed in which instructed trials and delayed trials were *both* classified in the same way, using an *a posteriori* criterion, based on RT. The experimental design deliberately over-emphasised the number of instructed rapid trials. To account for possible overestimation of differences due to an *a posteriori* criterion, subsampling methods were used (see methods, section 4.2). Results of this subsampling procedure are shown in figure 4.7. If participants had ignored the instruction signal, the mean proportion of correct instructed trials should have been around 50% in both conditions. Instead, the mean proportion of correct trials was 87.1±9% in the rapid condition; and 87.2±8% in the delay condition. This suggests that the *a priori* classification yielded similar populations than the *a posteriori* classification.

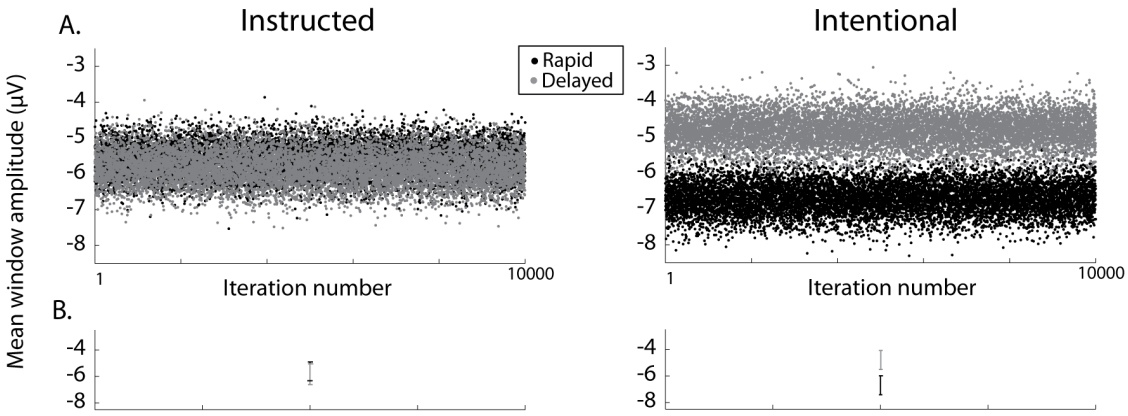


Figure 4.7: Bootstrapping procedure to resample instructed trials, accounting for differences in numbers of each trial type. This procedure allows instructed and intentional reactions to be classified based on reaction times. Results are shown for the interval of -150 to 0 ms prior to the onset of the visual cue, in electrode Cz **A.** The instructed rapid and delayed subsampled populations cannot be easily distinguished. In contrast, the intentional rapid and delayed subsample populations are clearly distinct. **B.** 95% Confidence intervals for instructed and intentional conditions. Note that while they are separate in the intentional conditions, they overlap in the instructed conditions. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

In the case of intentional conditions, the resampled data form two clearly distinct populations. One population of resampled trials presents slower, above-median RTs and is therefore classified as delayed/inhibited. Crucially, these results show that these trials were associated with lower prestimulus CNV amplitudes. A second population presented faster, below-median RTs. These trials were thus classified as rapid responses and showed higher prestimulus CNV amplitudes. The 95% confidence intervals for the two resampled populations do not overlap, replicating the finding of the main analysis. Prestimulus CNV amplitude differs before an intentional decision to respond rapidly or with a delay. In the instructed conditions, the resampled data do not form two clearly distinct populations and the 95% confidence intervals for prestimulus CNV amplitude show clear overlap between slower, above-median RTs classified as delayed/inhibited, and faster below-median RTs classified as rapid. The same resampling procedure was repeated for all three time windows, and is shown in figure 4.8.

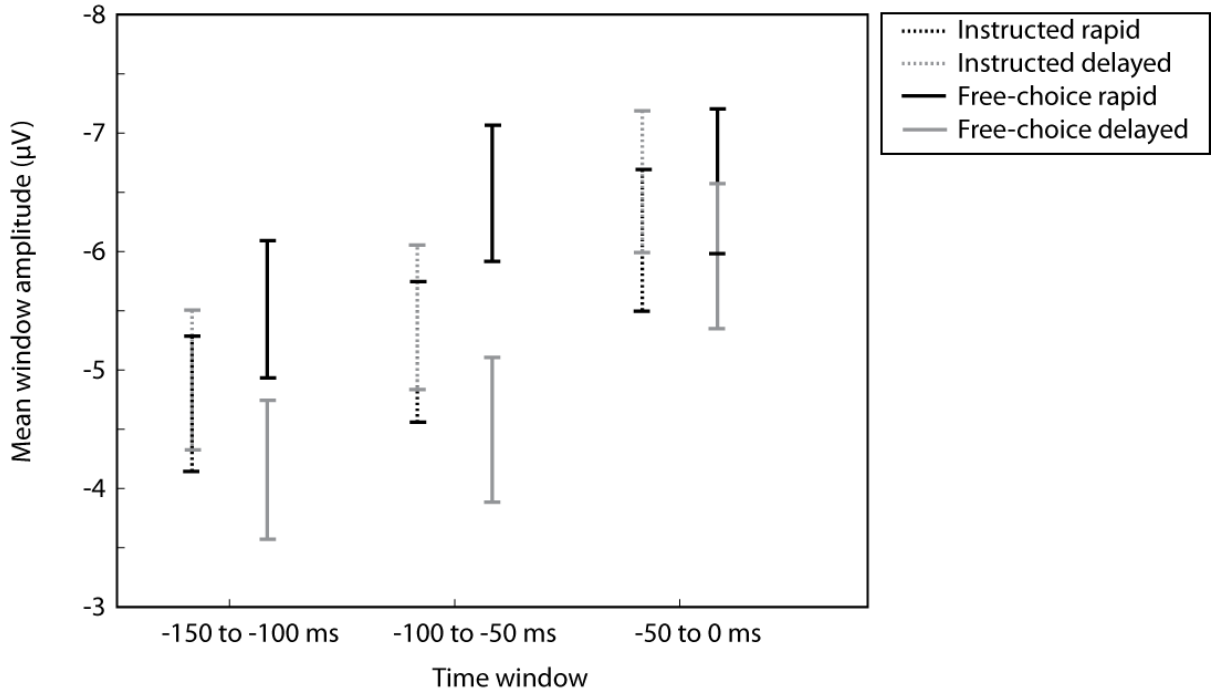


Figure 4.8 Results of the bootstrapping procedure in each 50 ms time window analyzed. Bars show the 95% confidence intervals of the mean ERP amplitude in Cz (as in panel B of figure 4.7). The first two time windows (-150 to -100 ms and -100 to -50 ms) show a significant source (instructed/free-choice) x outcome (rapid/delayed) interaction effect. Note that while the 95% confidence intervals do not overlap in the free-choice conditions, they do overlap in the instructed conditions. See table 4.1 for the corresponding results of statistical analysis. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

Lastly, the RT distribution for each participant was analyzed. CNV amplitude has been shown to be inversely related to RT (Hillyard, 1969). Therefore, the strong difference in the CNV amplitudes between the intentional conditions could in principle be a simple consequence of a strong separation between quick and delayed RTs. If this were the case, then RT distributions in the intentional conditions should be more strongly bimodal than instructed conditions. An established coefficient of bimodality *b* appropriate for large trial numbers (S. A. S. Institute, 1999):

$$b = \frac{(s^2 + 1)}{(k + 3)}$$

Where *s* and *k* are indexes of skewness and kurtosis respectively. The index of bimodality was in fact lower in the intentional than in the instructed condition (the instructed conditions showed a higher coefficient of bimodality than intentional conditions, mean (± SD) 0.44 ± 0.07 and 0.22 ± 0.05, respectively). These measures of bimodality were significantly different (t_{13}=9.7, $p < 0.01$). The difference remained significant when data was subsampled using the same procedures described above for the resampling of trials to calculate the CNV amplitudes. This provides further evidence against the possibility that CNV amplitude differences in the intentional conditions simply reflect stronger RT differences for intentional than instructed choices.

4. 4 Discussion

In this experiment, participants were instructed either to press rapidly or to inhibit and delay a key press; or they were free to choose between these two alternatives. The results show that the neural activity before the moment of decision to inhibit differed from that before a decision to act rapidly. When participants chose to respond rapidly on intentional trials, they did so on the basis of stronger preparatory activity *before* the moment of choice. Choosing to transiently inhibit and delay responding was associated with lower preparatory activity. This prestimulus influence on decision was unique to intentional trials, and was absent or reduced when participants were instructed to inhibit/delay. By definition, in the instructed condition, participants' behaviour was dictated by the instruction cue. Therefore the prestimulus CNV activity cannot predict instructed behaviour. Consequently, the instructed condition was used as a negative control, and differences in the prestimulus CNV trace between the two intentional conditions were sought.

Because different criteria were used to classify rapid and delayed trials for instructed and intentional trials, additional analyses in which instructed trials were also classified according to their RTs. The pattern of results remained the same. Therefore, a specific prestimulus CNV amplitude difference between rapid and

delayed actions was still present for intentional trials, but not for instructed trials, even when the number of trials was balanced across conditions, and classification criteria were chosen to distinguish rapid and delayed responses in a similar way for intentional and instructed decisions.

Could these results have occurred because of variations in general arousal level? Specifically, a participant who was mind-wandering or not engaged in the task might be expected to show low CNV amplitudes and long RTs (Tecce, 1972). Conversely, a high preceding level of arousal and engagement would be likely to produce a short RT. Thus, on an intentional trial, a prior state of high arousal would be likely to be classified as a decision to respond rapidly, even if no specific cognitive process of decision actually occurred. Similarly a low preceding level of arousal would be likely to be classified as a decision to transiently inhibiting responding. On this view, the relations between prior CNV activity and RT that were identified as decisions to act or inhibit might in fact be due to general arousal effects, rather than effects of prior neural activity on a specific cognitive decision process. However the variation in RT in the present data is much larger than that expected due to arousal effects alone. For example, Cheyne (Cheyne, Solman, Carriere, & Smilek, 2009) have described the "natural" fluctuations in RT in a go/nogo task (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). Their results show, for example, that trials preceding commission errors were on average 20 ms quicker than other trials. Conversely, trials preceding omission errors were on average 150 ms slower than the baseline. In the experiment reported here, the differences between rapid and delayed trials were of around 600 ms, much longer than delays explained by occasional inattention or "zoning out" episodes. These results suggest that the RT differences reflected outcomes of a specific decision process, and that this specific process was driven by neural precursor activity. This precursor activity may well have been in turn related to arousal, but the effects described here were clearly mediated by a specific *whether* decision process. This decision process occurred either based on external instruction, or on participants' intentional decisions. These results show that these intentional *whether* decisions in fact depended on preceding brain activity, *before* the cue requiring an intentional decision. The current state of the brain appears to influence the conscious decision to act or inhibit/delay, rather than vice versa.

Could participants actually have decided to inhibit/delay *before* the visual signal to choose? Two facts argue strongly against this potential predecision. First, frequent and rapid-response trials in the instructed condition were included to discourage such early predecision, participants were rewarded according to their RTs on these trials. Second, a 2x2 ANOVA revealed a main effect of source of decision, with intentional trials being 90 ms slower than instructed trials ($p=0.019$), consistent with a time-consuming decision stage occurring after the instruction, and comparable to RT costs of instructed choices (Hick, 1952).

Finally, to discourage stereotyped behaviour in the intentional trials (such as direct alternation between action outcomes), instructed and intentional trials were interleaved. In this way, a predecided strategy to maintain a stereotyped behaviour would have required higher working memory load. To check whether such predecision occurred, the distribution of the length of runs was examined (i.e., sequences of repeated action outcomes) for each participant. A distribution of runs strongly centred around a given number would have indicated a predecided strategy. However, no participant showed evidence for a stereotyped behaviour of this kind (see appendix A).

These data suggest that intentional decisions to inhibit/delay were made *after* the visual cue, but were strongly driven by antecedent, unconscious brain activity.

Limitations of this study

Several limitations should be considered when interpreting the results of this study. First, the sample size was relatively low, and inferences should therefore be tempered with caution. Nevertheless, the size of this study is comparable with other recent studies on prestimulus EEG activity (Britz & Michel, 2010; Busch & VanRullen, 2010; Mazaheri, DiQuattro, Bengson, & Geng, 2011).

In terms of design and data analysis, five important limitations should be taken into consideration. First, the factorial design was not perfectly balanced, as it included a relatively higher number of instructed quick trials as compared to instructed delayed

trials. Participants were rewarded on the basis of these rapid instructed trials only. At the end of each block they received a reward proportional to the number of rapid instructed trials that were faster than the average on the preceding block. Because only instructed rapid trials were rewarded, free and instructed conditions differed in terms of motivation. These differences in motivation may have influenced the way in which movements have been prepared or delayed. This imbalance in both trial numbers and reward was the result of a strategic decision to decrease the risk of participants predeciding before the trial whether to act or delay. By using instructed trials to ensure a motivation to prepare actions in advance, and rewarding participants accordingly, inhibition of an already-prepared action occurring before delayed responses could be assumed with more confidence.

Second, intentional trials were classified as rapid or delayed actions based on their reaction times. This approach has the advantage of not relying on subjective report, but only on objective behavioural measures. However, these objective measures may not provide a perfect classification of the participants' intentional decisions to respond rapidly or to delay their responses. Long RTs may be indicative of action inhibition, but may also arise for other reasons than inhibition, such as failures of attention, long decision times, etc. However, if this classification approach were simply imperfect, this would count against the probability of finding significant differences between trial types.

Arousal is one particular factor that might influence RT by affecting preparation. However, a *general* relation between arousal and RT would be presumably common to both instructed and intentional conditions. To explore the particular possibility of a role of arousal, resampling analyses were conducted by splitting the instructed data into rapid and delayed based on the median RT. In this way, had arousal been the *only* factor influencing the CNV amplitudes, then the instructed conditions would have shown two different populations in this resampling analysis. This was not the case. Instead, there was a *specific* relation between preparatory activity and an intentional decision to delay, with no such relation in instructed conditions. This cannot be explained by a general relation between arousal and RT without additional *ad hoc* assumptions.

Related to these considerations, it should be mentioned that the experimental design may have hindered the comparison between instructed and intentional trials. Because the SOA between the two visual cues (S1 and S2) was kept constant, intentional conditions may have been associated with weaker action preparation, as compared to the instructed conditions. This possible asymmetry, however, cannot alone explain the interaction pattern between intentional and instructed rapid and delayed conditions.

A third limitation of this study comes from the low spatial resolution of ERP (Luck, 2005). In particular, the differences in CNV amplitude which precede rapid vs. delayed intentional responses may have a subcortical source (Nagai et al., 2004) that cannot be measured at the scalp.

Fourth, this analysis may miss out some hemisphere-specific variations in preceding neural activity. Participants were asked to switch hands in every block, and then collapsed the ERPs obtained for the hemisphere contralateral and ipsilateral to the movement, regardless of the hand actually used for movement. However, the distribution of RPs in left and right hemisphere is known to differ e.g., (Wittmann, von Steinbüchel, & Szelag, 2001). Dirnberger et al [21] have shown that there are "atypical" trials in left hand key presstasks (but not in right hand key press tasks) with exceptionally early pre-movement activity. These atypical trials lead to RP amplitude distributions that violate the assumption of a Gaussian distribution, necessary for the parametric statistics used here. Because trials made with the right and left hand were collapsed, it is not clear how typical the RP amplitude distributions in this study are, and how valid the assumptions of normality are. However, the experimental design focussed on differences between intentional and instructed conditions, with equal numbers of right and left hand movements in each condition. Any bias introduced by hemispheric asymmetry should be equivalent in intentional and instructed conditions, and would therefore not influence the conclusions drawn here. Nevertheless, further control experiments could check for potential right-left asymmetries.

Also, and importantly, this experimental design cannot conclusively ensure that the prestimulus neural activity recorded is indeed unconscious. Stronger neural activity could in fact emerge as a *result*, and not a cause, of conscious intentions to act.

Notably however, chapter 7 in this thesis illustrates how problematic such distinctions between conscious and unconscious intentions may be. The alternative approach of converging behavioural evidence taken above may in this case provide stronger arguments that those that derive from directly probing the contents of conscious intentions.

Finally, as is common practice in paradigms involving intentional choices, participants were asked to try to balance their choices, and roughly choose to act rapidly in 50% of the intentional trials. This requisite for a roughly balanced behaviour may have encouraged participants to predecide in advance the sequence of intentional decision outcomes they would choose. This possibility was formally explored and no evidence was found to support non-random behaviour, but the possibility cannot be fully discarded.

Implications of this study

Intentional decisions about *what* action to make have been shown to be affected by subliminal primes (Schlaghecken & Eimer, 2004). In the same way, subliminal primes have been shown to modulate ERP components typically associated with inhibition in a go/nogo task (Hughes et al., 2009). This study did not use subliminal primes were not used to alter the preceding neural activity, but instead this study capitalized on the intrinsic variation in brain activity. It was assumed that the state of brain processing just before the instruction signal might influence a "free" decision about what actions when this was not specified in the instruction itself. These results strongly suggest that participants "freely decided" to respond quickly or delay their responses, depending on the degree of preparation within the cortical motor system immediately preceding the instruction to decide. These data can be parsimoniously explained by the suggestion that conscious intentional decisions to inhibit action may depend on the preceding state of the brain. Interestingly, the classic definition of voluntary actions involves contrasting them with instructed, stimulus-driven actions (R. E. Passingham et al., 2009). Volition thus amounts to "not externally generated" action. These cortical excitability measures would presumably satisfy this definition, since they correspond to fluctuations of internal signals. Links between free will and other internal neural signals have been proposed, notably the activity of the default mode network (I. Goldberg, Ullman, & Malach, 2008).

Antecedent brain activity was shown to precede subsequent conscious decisions about when to act (by about 700 ms -Libet et al., 1983-), or to be predictive of what action to perform (by several seconds -Soon et al., 2008-). EEG activity was also reported to precede intentional decisions to inhibit (Walsh et al., 2010). However, Walsh et al.'s results depend on interpreting *subjective* reports about time of intentional decisions, which remains controversial (Banks & Isham, 2009). Moreover, the experimental designs of those studies did not take the steps taken here to exclude advance pre-decision about whether to action or not. The task presented here was designed to constrain the *whether* decision to act or delay/inhibit to an identifiable point in time. This makes the finding of antecedent neural prediction of intentional decisions more striking, and may provide more convincing evidence for a form of neural determinism. In particular, these results show that antecedent brain activity influences intentional decisions. This was true even when the decision process was precisely defined in time, and when data analysis was based on objective behavioural criteria, rather than on subjective reports.

Importantly, these results also illustrate that unconscious brain activity significantly influences behaviour in situations where participants intentionally decide how to respond, yet there is no strong motivation to choose any one possible response alternative over the other. Preceding brain activity may have much less influence on behaviour when a clear instruction or strong internal motivation (such as a financial incentive) encourages choosing one response alternative over the other. In that case, any influence of preceding brain activity will be diluted or overridden to produce the "correct" response. On the other hand, cases of decision without clear instruction or strong internal motivation are particularly important, because they are the focus of debates about "endogenous" decisions, and more generally about "free will".

The main argument is as follows: Libet et al, (1983) had suggested that decisions to inhibit action have an important role in freedom of will, because, he argued, they do not have any obvious unconscious neural precursors. In Libet's view, this makes them crucially different from decisions to act, for which, he claimed, there *is* a clear unconscious precursor. Libet's dualistic notion of "free won't" has been criticised on theoretical grounds. However, a stronger rejection of "free won't" could come from

actually showing that the decision to act or not is driven by a preceding, presumably unconscious neural activity. The present results identify, for the first time, a candidate unconscious precursor of the decision to inhibit action. These results count as evidence against Libet's view that the decision to inhibit action may involve a form of uncaused conscious causation.

4. 5 Conclusion

Neuroscience cannot straightforwardly accommodate a concept of "conscious free will", independent of brain activity (Haggard, 2008). However, the belief that humans have free will is fundamental to human society (Nichols, 2011). This belief has profound top-down effects on cognition (Vohs & Schooler, 2008) and even on brain activity itself (Rigoni, Kühn, Sartori, & Brass, 2011). The dualistic view that decisions to inhibit reflect a special "conscious veto" or "free won't" mechanism (Libet, 1985) is scientifically unwarranted. Instead, conscious decisions to check and delay actions may themselves be consequences of specific brain mechanisms linked to action preparation and action monitoring (Brass & Haggard, 2007). Recent neuroscientific studies have strongly questioned the concept of free will, but have had difficulty addressing the alternative concept of free won't, largely because of the absence of behavioural markers of inhibition. These results suggest that an important aspect of "free" decisions to inhibit can be explained without recourse to an endogenous, "uncaused" process: the cause of "free decisions" may at least in part, be simply the background stochastic fluctuations of cortical excitability. These results suggest that free won't may be no more free than free will.

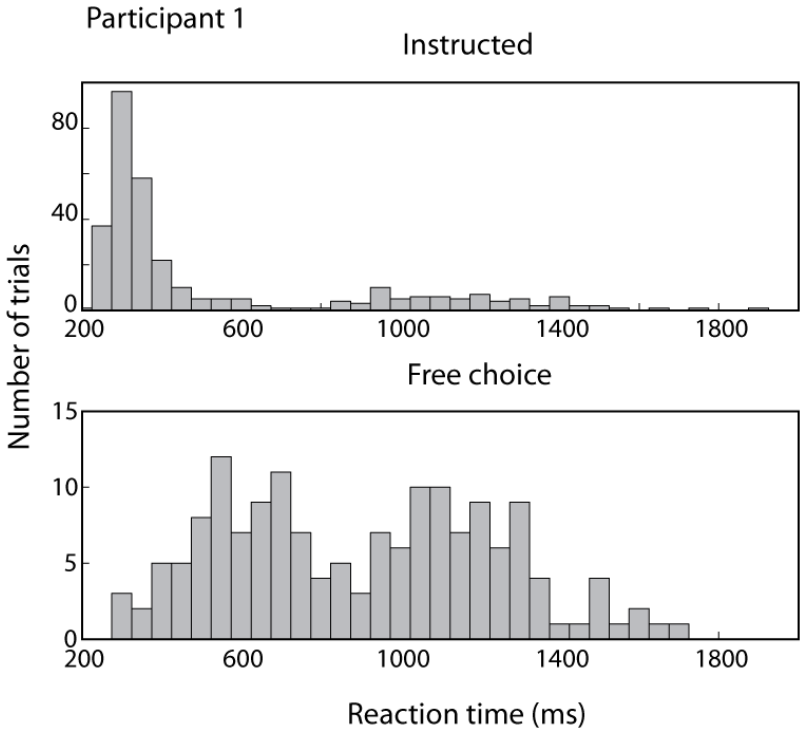
4. 6 Appendix

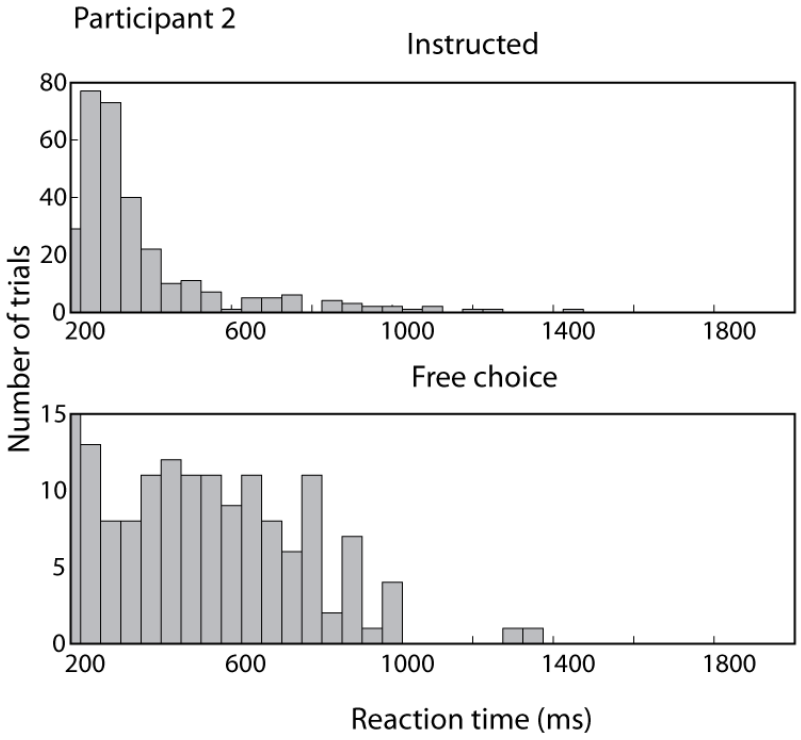
Figure A1 shows the individual RT distributions. For each participant, top histograms show RTs in the instructed conditions. Bottom histograms show RTs in the free choice conditions.

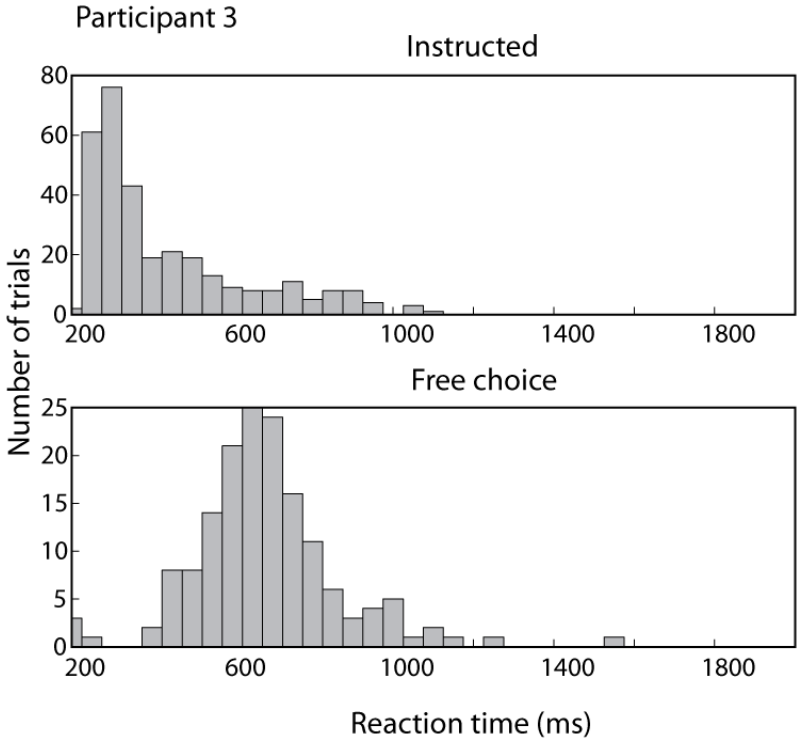
To encourage preparation, the experimental design included a high proportion of instructed quick trials (see Methods). The RT distributions reflect this imbalance. A bimodal RT distribution is nevertheless evident in most participants for the instructed conditions.

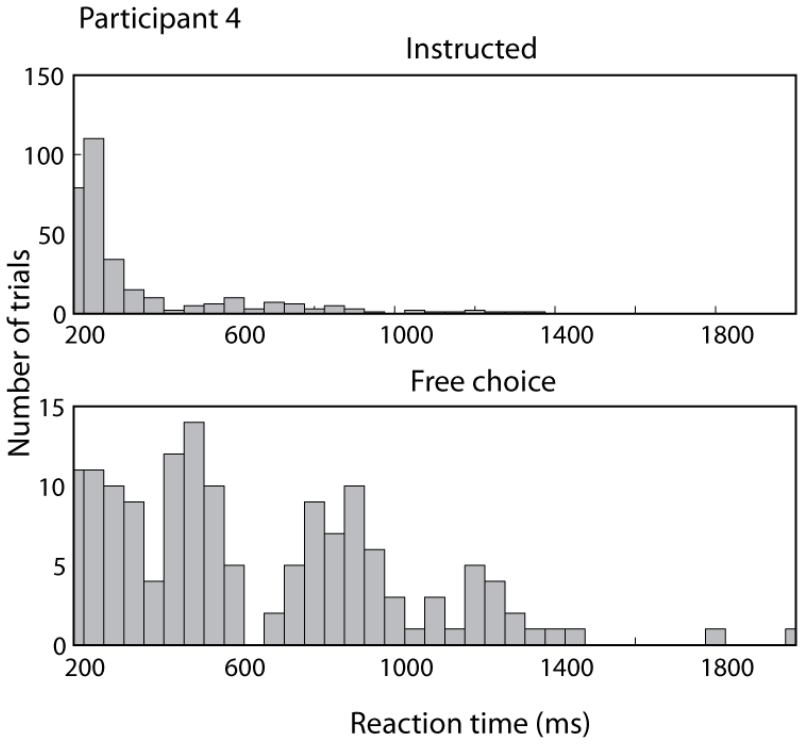
Free choice conditions show more uniform distributions, with smaller values of bimodality, see main text. Importantly, this argues against the possibility that the observed ERP effects are simply due to differences in RTs. In particular, a more uniform distribution in the free choice conditions would have lead to smaller (and not larger, as it was the case) ERP differences between quick and delayed conditions.

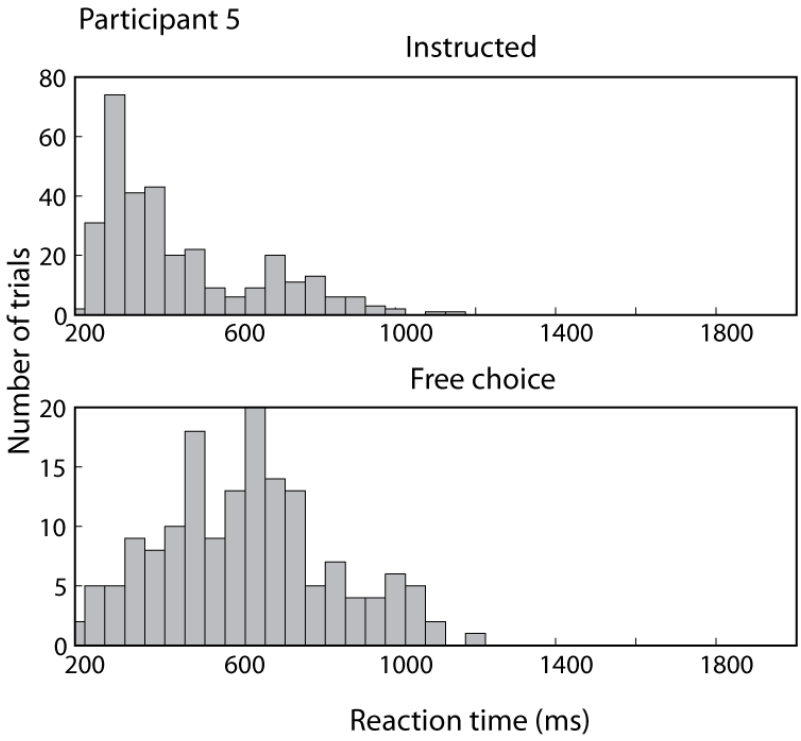
Fig A1: Individual RT distributions for each participant. Time bins represent 50 ms. Note that y axis values are different for each participant. Adapted from Filevich et al (in Press), published under Creative Commons Attribution License.

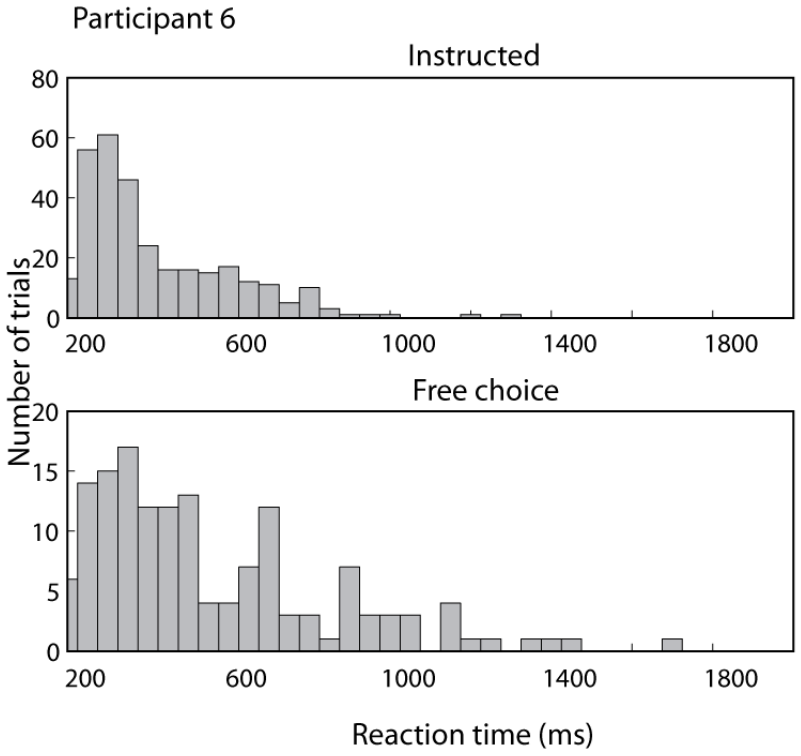


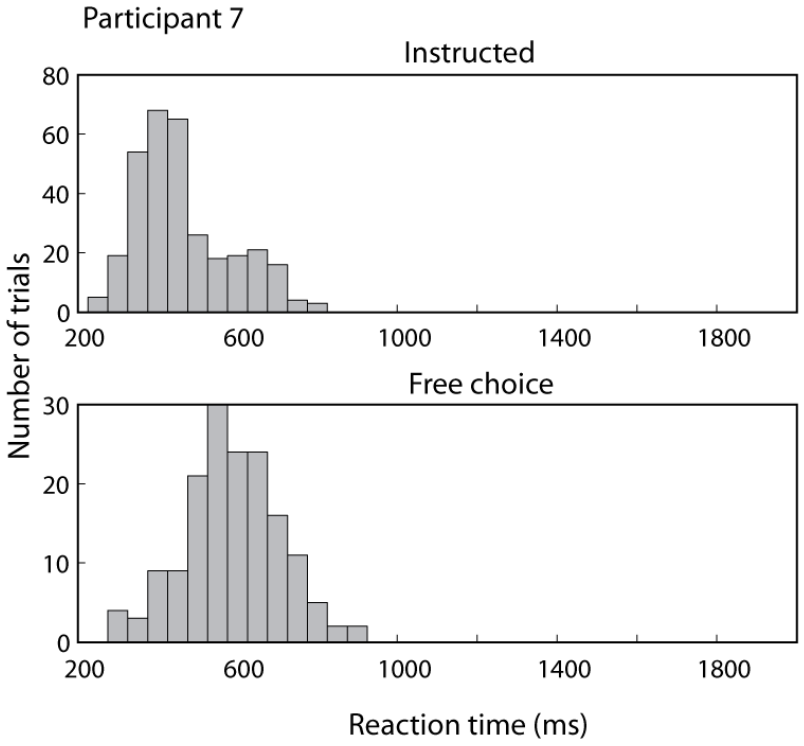

Participant 2
Instructed
Number of trials
80
60
40
20
0
200
600
1000
1400
1800
Free choice
15
10
5
0
200
600
1000
1400
1800
Reaction time (ms)

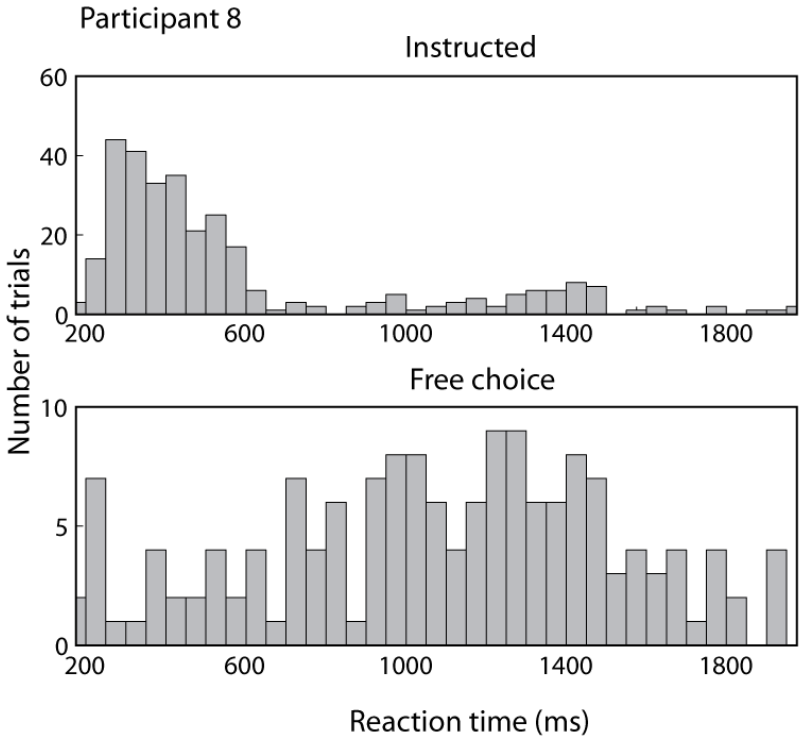

Participant 3
Instructed
Number of trials
80
60
40
20
0
200
600
1000
1400
1800
Free choice
25
20
15
10
5
0
200
600
1000
1400
1800
Reaction time (ms)

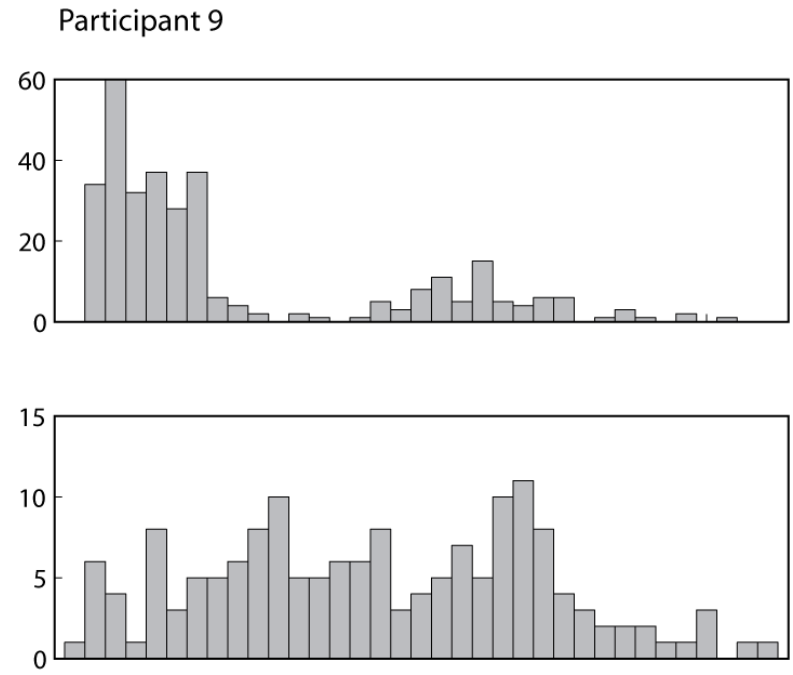

Participant 4
Instructed
150
100
50
0
200
600
1000
1400
1800
Free choice
15
10
5
0
200
600
1000
1400
1800
Number of trials
Reaction time (ms)

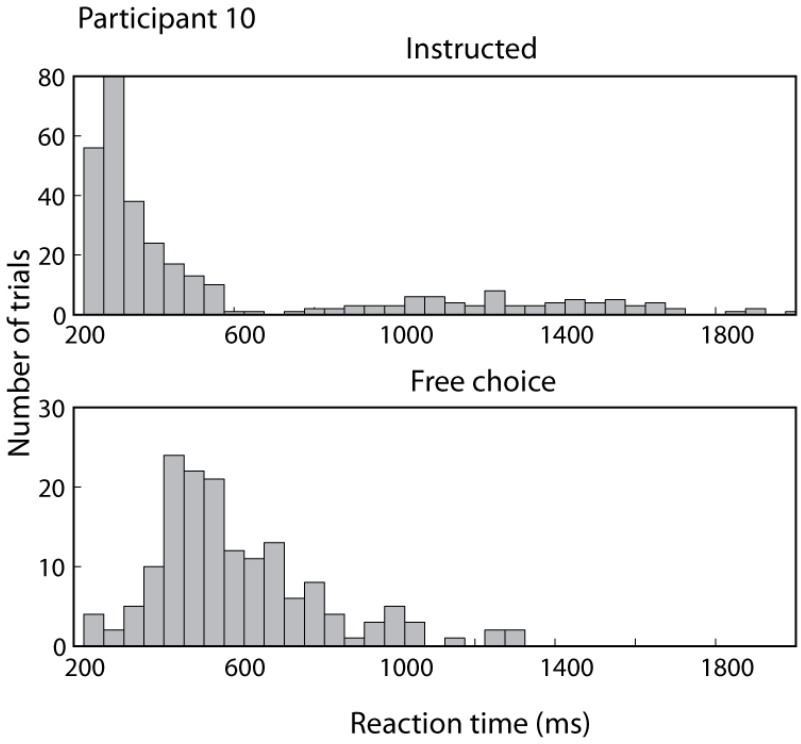

Participant 5
Instructed
80
60
40
20
0
200
600
1000
1400
1800
Free choice
20
15
10
5
0
200
600
1000
1400
1800
Number of trials
Reaction time (ms)

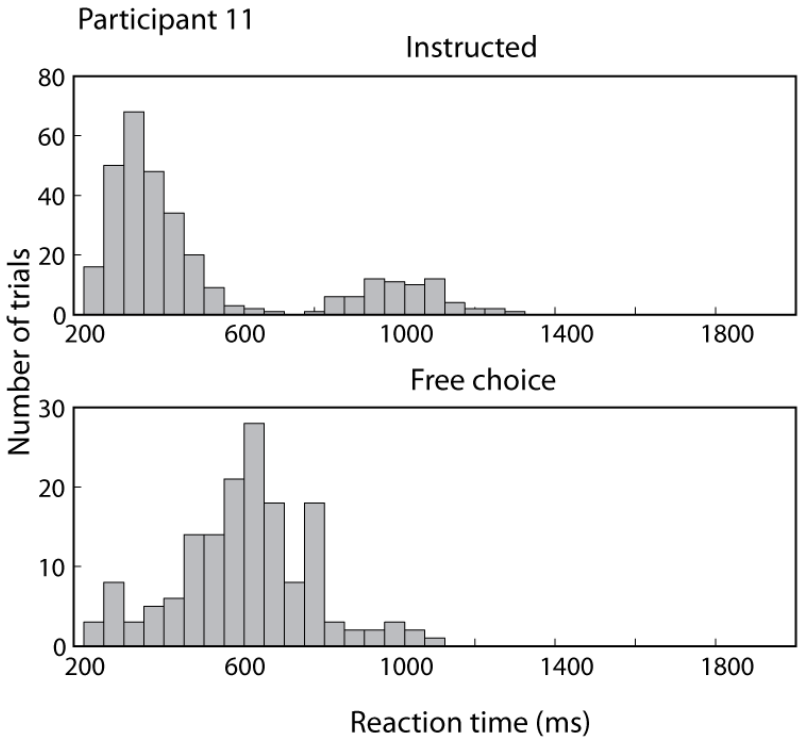

Participant 6
Instructed
Free choice
Number of trials
Reaction time (ms)
0
20
40
60
80
5
10
15
200
600
1000
1400
1800

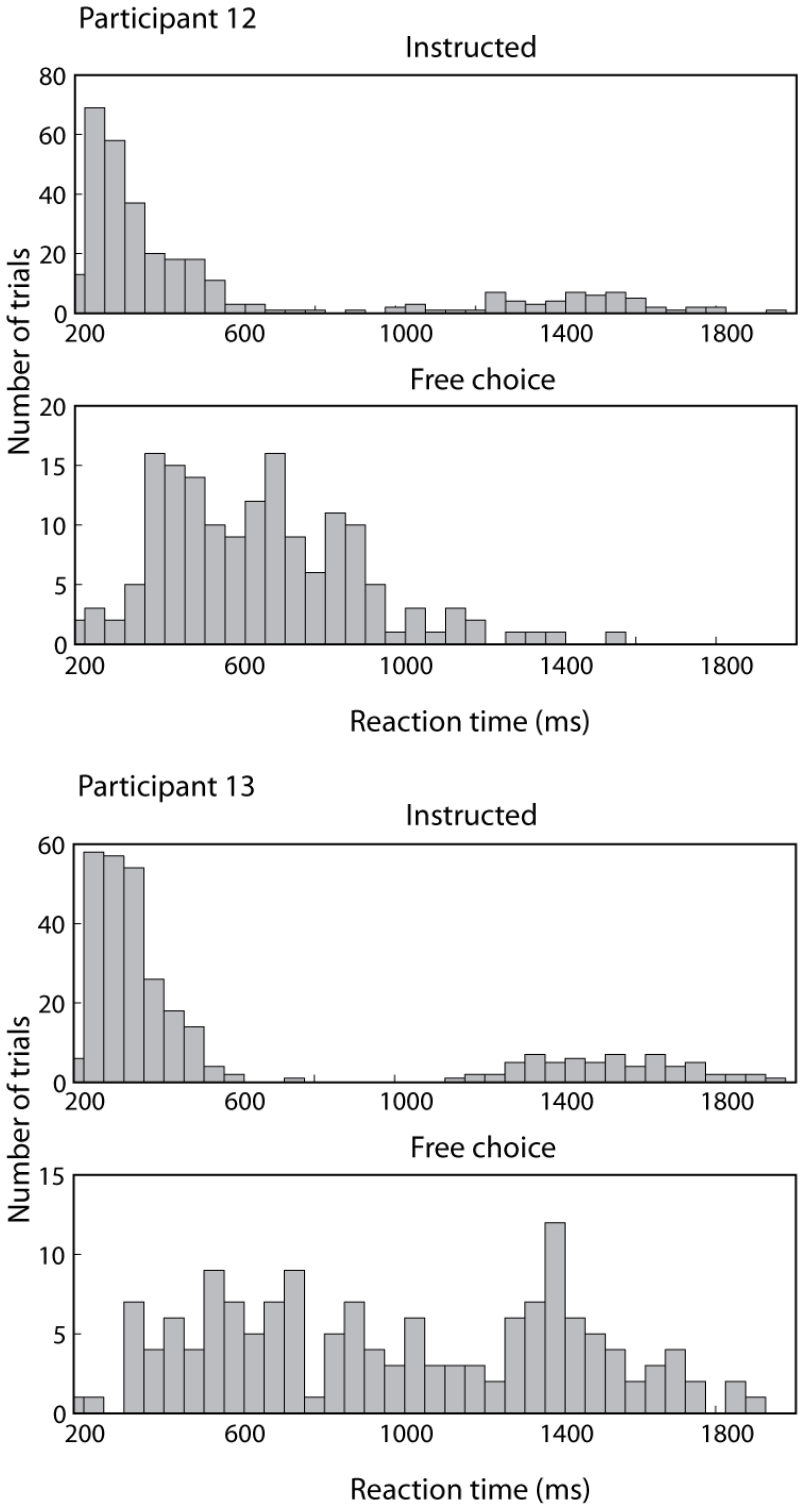

Participant 7
Instructed
Free choice
Number of trials
Reaction time (ms)
0
10
20
30
40
60
80
200
600
1000
1400
1800

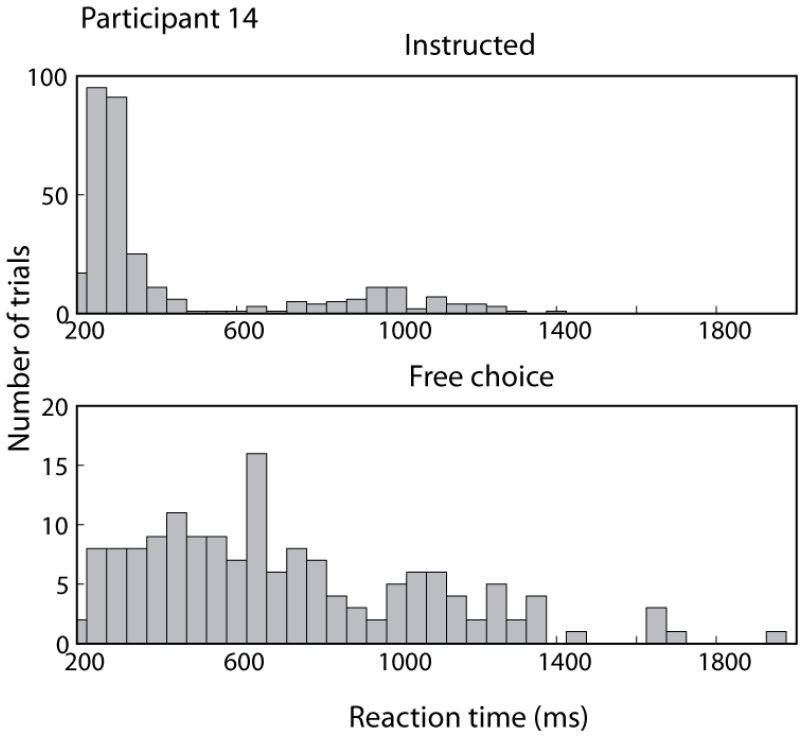

Participant 8
Instructed
60
40
20
0
200
600
1000
1400
1800
Number of trials
Free choice
10
5
0
200
600
1000
1400
1800
Reaction time (ms)


Participant 9
60
40
20
0
15
10
5
0


Participant 10
Instructed
Free choice
Number of trials
Reaction time (ms)
0
20
40
60
80
10
30
200
600
1000
1400
1800


Participant 11
Instructed
Free choice
Number of trials
Reaction time (ms)
0
20
40
60
80
10
30
200
600
1000
1400
1800


Participant 12
Instructed
Free choice
Number of trials
Reaction time (ms)
Participant 13
Instructed
Free choice
Number of trials
Reaction time (ms)


Participant 14
Instructed
100
50
0
200
600
1000
1400
1800
Free choice
20
15
10
5
0
200
600
1000
1400
1800
Number of trials
Reaction time (ms)

Chapter 5 Relative processing of task-irrelevant stimuli

This study investigated whether intentional actions may be easier to modify than instructed actions, by measuring susceptibility to external distractors. ERP amplitudes to task-relevant and task-irrelevant stimuli were compared situations of instructed or intentional action preparation. Intentional action was associated with less efficient suppression of task-irrelevant information than instructed action. Intentional action may therefore being more susceptible to distraction.

5. 1 Introduction

One of the working hypotheses of this thesis is that intentional decisions for action are weak as compared to instructed decisions for action. This hypothesis contrasts with evidence from previous studies comparing readiness potential (RP) or lateralized readiness potential (LRP) amplitudes between situations of intentional and instructed actions. Typically, these studies have found *stronger* RP amplitudes for intentional actions e.g. (Praamstra et al., 1995), and no differences in LRP amplitude between conditions e.g., (Waszak et al., 2005). One possible explanation that may reconcile these apparently contradicting results is that the stronger RPs observed are related to movement expectancy and anticipation, and consequently have a weak functional relevance for movement preparation.

One previous study investigated potential *functional* differences between instructed and intentional decisions, by measuring their relative flexibility (Fleming, Mars, Gladwin, & Haggard, 2009). In an S1-S2 task, Fleming *et al* asked participants to prepare to make a right or left key press following an initial visual cue (S1). Two types of S1 cues were possible. Instructed S1 cues specified the hand with which participants had to prepare to make the action. Intentional S1 cues, on the other hand, allowed participants to freely choose between right- and left-hand key presses. After a 1.5 s delay, a second cue (S2) followed S1. S2 could be either a "stay" or a "change" cue. S2 *stay* cues meant that participants simply had to make the action they had prepared following S1. S2 *change* cues, on the other hand, instructed participants to change their action plans. That is, participants were required to make a key press with their left hand if they had prepared a right-hand key press and vice versa.

Fleming *et al* then compared the neural responses to S2 between intentional and instructed conditions. They found that ERPs evoked by S2 change cues showed greater amplitudes in instructed trials as compared to intentional trials. Thus, the authors suggested that changing instructed action plans required more neural effort than intentional action plans, and must have therefore a more robust and less flexible underlying neural code.

Such indirect measures of action preparation can reveal differences between intentional and instructed decisions for action. This study explored another way of measuring strength of decisions, namely their resistance to external distraction. In particular, intentional action preparation may be more susceptible to interference from external, task-irrelevant stimuli than instructed action preparation.

P300 as a measure of information processing

As in Fleming *et al* (Fleming et al., 2009), numerous studies have measured the amplitude of the P300 component, or of one of its subcomponents as a proxy for the strength of processing of external stimuli in general, and of task-irrelevant stimuli in particular (Bledowski et al., 2004; Gumenyuk et al., 2001; Polich & Ochoa, 2004; Richard Clark, McFarlane, Weber, & Battersby, 1996). Both task-relevant and task-irrelevant stimuli typically elicit a large positive component in the EEG signal, termed P300, peaking at around 300-500 ms after stimulus onset. The analysis of neural activity following task-relevant and task-irrelevant stimuli provides a means to measure the automatic processing of external stimuli. For example, Vallessi and Stuss (Vallesi & Stuss, 2010) tested younger and older participants in a modified go/nogo task. The authors measured LRP amplitudes in response to nogo stimuli. Although these task-irrelevant nogo stimuli did not require a motor response, older adults showed significant LRPs developing in nogo trials, consistent with the automatic preparation of a response, despite no need form an overt motor action. The authors therefore suggested that older adults showed less efficient mechanism of suppression of the processing of task-irrelevant stimuli.

To test the hypothesis of differential strengths of instructed and intentional decisions, this study capitalized on the methods available to assess the automatic processing of task-irrelevant stimuli. The amplitudes of the neural response to task-relevant and task-irrelevant stimuli were compared between situations of instructed and intentional action preparation. The experimental design followed closely that of Fleming *et al* (Fleming et al., 2009). Here, their original design was extended to include task-irrelevant stimuli instead of the otherwise always relevant *stay*/change S2 cues.

The present task required participants to prepare an action following an S1 cue, and maintain action preparation until an S2 cue informed whether these action plans should be changed or not. Some S2 cues were task-relevant, and had to be followed, whereas other S2 cues were irrelevant and had to be ignored. Because quick key presses were encouraged, optimal performance in this task required a suppression of the perceptual and cognitive processing of task-irrelevant stimuli. The hypothesis was as follows: if instructed and intentional action decisions differ in the relative strengths of their underlying neural codes, then task-irrelevant stimuli presented in the context of strong instructed action preparation will be efficiently suppressed, and processed less deeply than task-relevant stimuli. On the contrary, task-irrelevant stimuli presented in situations of intentional action preparation will only show intermediate levels of perceptual suppression, due to the intermediate levels of action preparation. The processing of task-irrelevant stimuli will therefore be similar to that of task-relevant stimuli

5. 2 Methods

Participants

Sixteen naïve participants (nine females, mean age ± SD, 26 ± 8) participated in this study. One participant was left-handed. Procedures were approved by the UCL research ethics committee and were in accordance with the principles of the Declaration of Helsinki. In total, 4 participants were excluded without further analysis. 2 participants blinked excessively and the data from 2 participants could not be recorded due to technical problems, yielding a total of 12 participants included in the analyses.

Experimental task

Participants sat in a dimly lit and quiet room, 60 cm away from a stimuli display screen. The experiment was divided in 6 blocks of 128 trials each. Trial order was

randomized for each participant. The experiment lasted approximately 90 min, and participants got familiarized with the task during a short practice session before starting with the experiment.

Each trial was structured as follows (see figure 5.1): a white fixation cross appeared at the centre of a black screen for a variable duration of between 500 and 1200 ms. Two visual stimuli (S1 and S2) were presented for 250 ms each, with a fixation cross period of 700 ms between them. Participants were asked to respond with their right or left index fingers following S2 by making a key press on a standard computer keyboard. Participants had a time limit of 1250 ms to make their response. The inter trial interval was 800 ms.

Three main trial categories were included in the task. These were S1-S2 trials (together, 63%), short trials (15%) and baseline trials (22%). The main experimental conditions (S1-S2 trials) consisted in turn of 8 possible variations (see figure 5.1). These variations were a combination of 3 independent factors, namely source of decision (instructed/intentional, indicated by the S1 cue); and relevance (task-relevant/task-irrelevant) and outcome (*swap*/*stay*), both given by the S2 cue.

S1-S2 and short trials could be either instructed or intentional. In instructed trials, S1 contained arrows pointing either to the right (>>) or to the left (<<), and participants were asked to prepare a key press with their corresponding hand. In intentional trials, a double-headed arrow (<>) was presented, and participants were asked to prepare a key press with a hand of their choosing. It was emphasized to participants that they should commit to a given hand, and try to avoid producing stereotypical sequences of responses. For the purposes of data analysis, right and left key presses were collapsed into a single condition.

700 ms after the offset of S1, one of four possible S2 cues was presented. S2 provided a second instruction that that could change the required key press. S2 cues displayed the words “swap” or “stay” written in lower- or upper-case. Upper- or lower-case words could correspond to either task-relevant or task-irrelevant stimuli. This mapping was counterbalanced across subjects, and reversed once for each participant halfway through the experiment. Task-relevant stimuli were to be followed, whereas task-irrelevant stimuli were to be ignored. Consequently, S2

conveyed one of two possible messages. A task-relevant S2 cue would either instruct participants to reverse (*swap*) or maintain (*stay*) their action plans. Instead, participants had to ignore task-irrelevant S2 cues, and execute their action plans prepared following S1. The 2x2 factorial manipulation for S2 (relevance x outcome) was crossed over with the two possibilities for S1 (factor of relevance), yielding a 2x2x2 factorial design. It is important to note, however, that only one (task-relevant *swap*) of the four possible S2 cues effectively required a change in the prepared action plan.

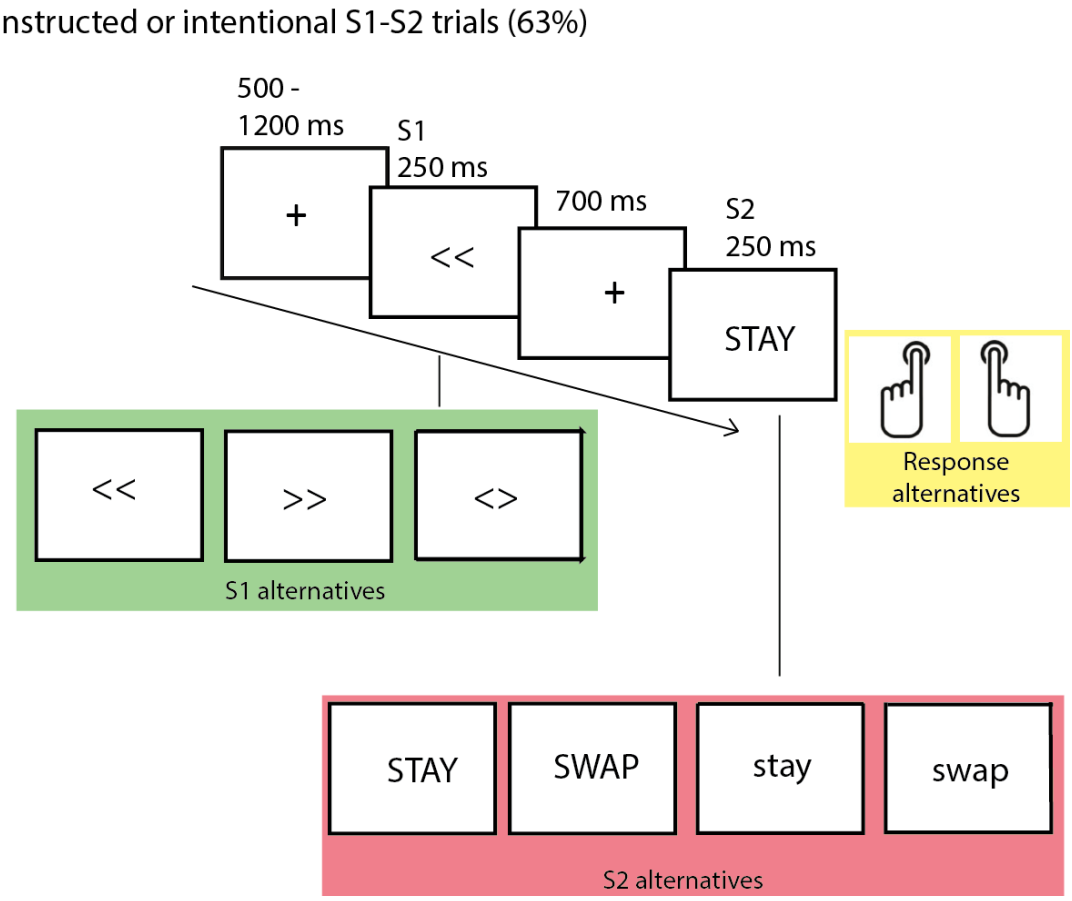


Figure 5.1: Main experimental conditions (S1-S2 trials). **V**isual stimuli S1 and S2 were presented with a fixed (700 ms) interval between them. Participants had to prepare a key press according to the S1 cue. S1 would either instruct participants to prepare a right (>>) or left (<<) key press, or allow them to intentionally decide which action to prepare (<>). Participants were asked to maintain action preparation between S1 and S2, and make a speeded key press following S2. S2 cues could require participants to continue with their action plans ("*stay*" cues) or to inhibit their prepared action and make a key press with the alternative, non-prepared hand ("swap" cues). S2 could either be a task-relevant or task-irrelevant cue (upper or lowercase, mapping counterbalanced across participants). Task-relevant S2s had to be followed, whereas task-irrelevant S2s had to be ignored.

The experimental stimuli chosen by Fleming *et al* for S2 were square and diamond-shaped visual stimuli that had been arbitrarily paired with the *stay*/*swap* instructions. In contrast, in the present experiment, word stimuli were preferred, for two main reasons. First, pilot data suggested that the pairing of four arbitrary stimuli with their corresponding instructions was relatively difficult and confusing for participants. More importantly, S2 cues were designed to be both easy to distinguish and difficult to ignore completely. Task-relevant and task-irrelevant stimuli were therefore different in their surface form, making them easy to distinguish without attending to their semantic content. At the same time, they were made difficult to ignore completely by using word stimuli, that could trigger automatic semantic processing (Stroop, 1935).

To encourage action preparation, the paradigm included short trials (15% of total trials). In these trials, the sign 'XXXX' appeared shortly after S1 (see figure 5.2 A). The short cue latency was randomly sampled from 5 possible latencies, between 200 and 600 ms. Participants were asked to make a key press promptly following a short cue, executing the prepared action in S1. Participants received a monetary reward of an amount proportional to the number of trials with an RT shorter than their own average in the previous block. Importantly, short trials were indistinguishable from S1-S2 trials, until the time at which S2 or the short cue appeared. Consequently, rewarding quick key presses in short trials was expected to encourage action preparation in both short and S1-S2 trials. Feedback was given when incorrect responses were made in instructed trials.

Finally, to confirm that participants had indeed prepared their actions following S1, baseline trials were included in the task (28% of total trials, see figure 5.2 B). In these trials S1 showed a question mark and was followed by a right (>>), left (<<), or double-headed (<>) arrow at the time of S2. Participants were asked to make a quick key press in the direction indicated by S2.

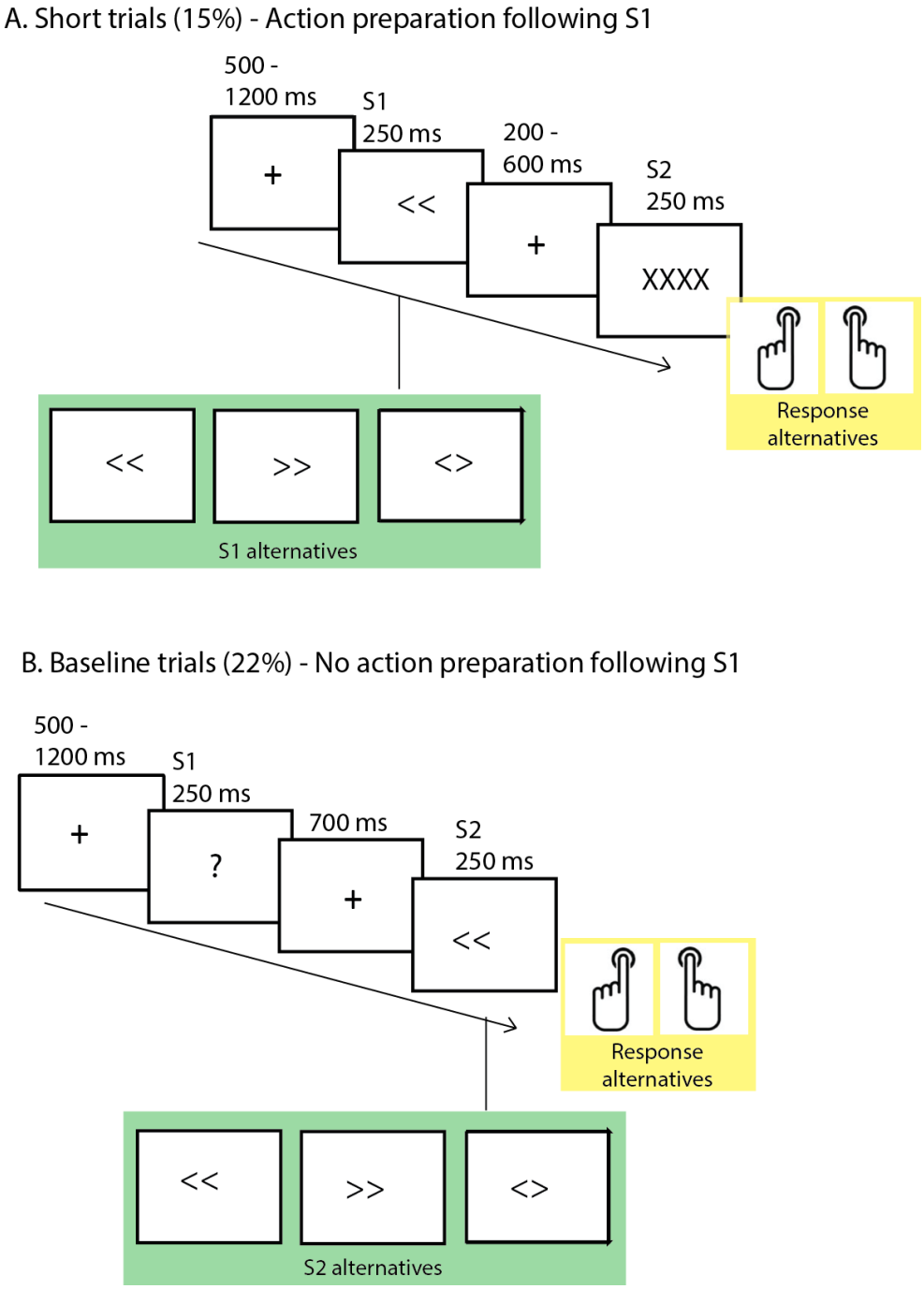


Figure 5.2 Short and baseline conditions. Two conditions were included to address whether participants prepared their actions after S1. **A.** Short trials were indistinguishable from the S1-S2 trials at the time of S1. However in contrast to an S1-S2 trial, short trials included a go signal ('XXXX') appearing at a variable interval before the expected time of S2. To encourage action preparation following S1, participants were asked to execute the key press they had prepared following S1, and were given a monetary reward proportional to the number of quick reaction times. **B.** In baseline trials, a question mark appeared at the time of S1, so participants could not prepare their action. S2 then either instructed participants to make a key press with their right (>>) or left (<<) hand, or internally decide which hand to use (<>). Action preparation was therefore not possible in baseline trials, and served as a measure of RTs in the absence of any preceding action preparation.

Electrophysiological Recordings and data analysis

EEG data were recorded with a SynAmps amplifier system and Scan 4.3 software (Neuroscan, El Paso, TX) from fourteen Ag/AgCl scalp electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, P3, Pz, P4, O1, O2, according to the 10-20 system). Electrodes were referenced to AFz online and the ground electrode was placed on the nose. Activity from left and right mastoids was recorded. Vertical and horizontal electrooculographic (EOG) activity was recorded from bipolar electrodes positioned above and below the right eye and on the outer canthi, respectively. Electromyographic activity was recorded from a pair of surface electrodes placed on the *flexor pollicis brevis* and the *adductor pollicis* muscle of each hand. Impedances were kept below 5 KΩ for all electrodes. EEG signals were amplified and digitized at 500 Hz; and recorded with an online notch filter between 45 and 55 Hz.

EEG data were analyzed with EEGLAB software (Delorme and Makeig, 2004). Data were first re-referenced to the linked mastoids. Data were digitally band-pass filtered between 0.05 and 30 Hz. All trials were collapsed irrespective of the hand used for the keypress. Consequently, to prevent the averaging out of lateralized potentials, the electrodes ipsilateral and contralateral to the movement were inverted for those trials that resulted in a left hand key press. Epochs were time locked to the onset of either the visual cues (S1 or S2) or to the time of the key press. Trials time locked to the visual stimuli were baselined to the period of 150 ms immediately prior to cue onset. Incorrect instructed trials were excluded from the analysis. Incorrect intentional trials could not be identified, so were not discarded (but see below). Further, to avoid artefacts due to eye blinks, trials were discarded if the bipolar recording of EOG exceeded +/- 80 μV at any point during the epoch.

A grand average was obtained for all 8 S1-S2 conditions. Time windows and electrodes of interest were defined on the basis of the peak latencies and topographical distributions of the grand average of all conditions together. Means across entire time windows were obtained, and entered into statistical analyses.

Evidence for lateralization of a prepared response during the period between S1 and S2 was sought. The standard measure of lateralization, namely LRP amplitude, could

not be obtained here. To compute LRPs, the EEG signal should be time locked to the time key press. In this task the definition of LRP epochs would pose a problem because the strong neural responses to the S2 cues, prior to actions, fell at variable times relative to the time of key press, contaminating the signal and hindering the LRP estimation. The critical frequency components of LRP do not overlap with those of P300 (Demiralp, Ademoglu, Comerchero, & Polich, 2001; Demiralp, Ademoglu, Schürmann, Basar-Eroglu, & Basar, 1999). Therefore, the approach adopted by Fleming *et al* was followed, and the motor-related amplitude asymmetry (MRAA) was computed as a measure of lateralization of the prepared response. The EEG signal was time locked to the time of S2, and epochs were defined between -1500 and 100 ms around the time of S2. The μ-band (9-13Hz) power was calculated for each trial by Morlet wavelet convolution. The average μ-band power for each participant and each condition in the C electrode ipsilateral to movement was subtracted from the average power in C electrode contralateral to movement. Finally, to obtain the MRAA measure, the contra - ipsilateral μ-band power difference was normalized for each participant to the average power for the whole epoch and across both contra and ipsilateral electrodes (Gladwin, Lindsen, & de Jong, 2006).

5. 3 Results

Participants made few errors across all instructed S1-S2 conditions (7.9 ± 6.76%). Incorrect trials could not be objectively identified in intentional conditions, because the covert action selection following S1 was not available to an external observer.

Behavioural evidence for action preparation

This experiment focussed on the modulatory effects of the state of action preparation following S1, on the neural processing of S2. Therefore to validly compare instructed and intentional trials, it was crucial that comparable levels of directional action preparation occurred following S1, in both intentional and instructed trials. Because S1-S2 trials were indistinguishable from short trials at the time of S1 (see

methods), evidence for action preparation following S1 in short trials constitutes indirect evidence for action preparation following S1 in S1-S2 trials. Therefore, to verify that action preparation indeed occurred following S1, RTs associated with short and baseline trials were compared. Action preparation at the time of S1 was impossible in baseline trials, but was encouraged in short trials. To search for evidence for action preparation following S1, RTs were compared in a 2x2 ANOVA with the factors of condition (short/baseline) and source (instructed/intentional choice) (see figure 5.3). Results revealed a significant main effect of source ($F_{1,11}$=16.70, p=0.002), consistent with an increased RT cost of decision (H. Lau, Rogers, Ramnani, et al., 2004; van Eimeren et al., 2006). There was also a significant main effect of condition ($F_{1,11}$=5.15, p=0.044), suggesting that there was indeed action preparation after S1, in both intentional and instructed trials. Further, there was a significant interaction effect ($F_{1,11}$=12.02, p=0.005). Follow-up t-tests revealed that whereas RTs in the instructed conditions did not differ significantly (t_{11}=1.68, p=0.120), RTs in intentional choice were in fact significantly different (t_{11}=2.71, p=0.020). Quicker RTs were associated with in intentional short trials as compared to those in intentional baseline trials. Although the difference in instructed trials was not significant, this result importantly confirms that action preparation occurred following S1 in the crucial intentional conditions.

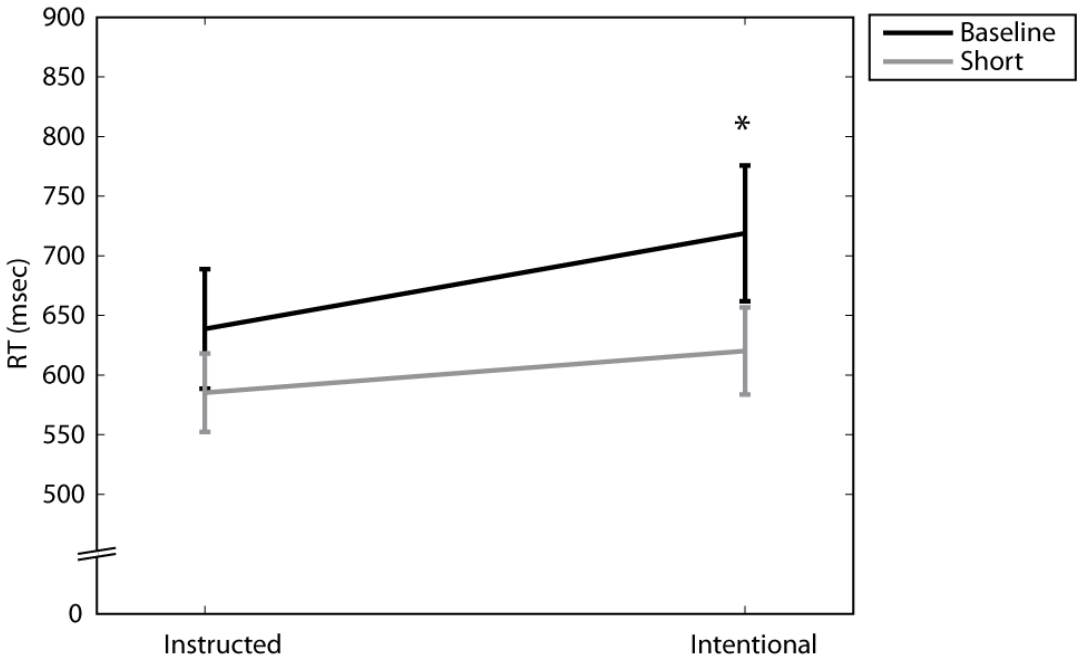


Figure 5.3 Evidence for action preparation following S1. Mean reaction times for short and baseline trials were compared. Action preparation was encouraged in short trials but impossible in baseline trials. Accordingly, RTs were quicker in short trials as compared to baseline trials, in both instructed and intentional conditions. The differences were stronger for intentional conditions. Error bars show standard error of the mean. ** $p<0.05$

The mean RTs were then compared for the main 8 experimental conditions in a 2x2x2 ANOVA with the factors of source (instructed/intentional), relevance (task-relevant/task-irrelevant) and outcome (*stay*/*swap*) (see figure 5.4). There was no significant main effect of source ($F_{1,11}=0.34$, $p=0.571$) but a significant main effect of outcome ($F_{1,11}=21.49$, $p<0.001$). An expected RT cost for *swap* trials was consistent with the main effect of outcome. If participants were indeed changing action plans at the time of S2 in *swap* trials, this effect should have been present in both instructed and intentional trials. There was, however, an additional significant interaction effect between source and outcome ($F_{1,11}=6.044$, $p=0.032$).

Because the action plans following S1 were private to participants, successful switching of action plans could not be objectively evaluated in intentional trials. Therefore, behavioural evidence was sought to ensure that switching effectively occurred in intentional trials. More specifically, the source x outcome interaction was explored to ensure that *swap* RT costs were present in intentional *swap* trials. Because there was no significant three-way interaction, the data were collapsed

across the two relevance levels. Paired t-tests revealed significant *swap* RT costs for both instructed (t_{11}=-6.14, p<0.001) and intentional trials (t_{11}=-2.24, p=0.046). Therefore, although the effect was more pronounced in instructed trials, intentional trials also presented a *swap* RT cost, suggesting true commitment to decisions in intentional trials.

Effects of task-irrelevant stimuli

The main hypothesis in this study was that intentional trials are more susceptible to the effect of task-irrelevant stimuli than instructed trials. If this is true, then RTs differences between task-irrelevant and task-relevant S2 cues should have been smaller for intentional trials as compared to instructed trials. Mean RTs for each conditions are shown in figure 5.4 The ANOVA on the mean RTs revealed a significant main effect of relevance ($F_{1,11}$=5.85, p=0.034), indicating that task-relevant stimuli demanded more cognitive processing. Importantly, however, the relevance of the S2 cue did not interact with the source of decision ($F_{1,11}$=0.14, p=0.712). Also, there was a main effect of outcome ($F_{1,11}$=21.48, p<0.001), suggesting that the processing of the *swap* cue itself led to increased RTs. However, there was no significant relevance x outcome interaction ($F_{1,11}$=3.16, p=0.103) and no significant three-way interaction ($F_{1,11}$=0.11, p=0.748). The absence of a significant interaction effect between outcome and other factors indicated that this RT increase was not exclusive to the actual need for action plan switching. Instead, increased RTs following *swap* cues may reflect increased perceptual processing of the stimulus, or perhaps automatic transient stopping of prepared action plans (Verbruggen & Logan, 2008b).

In sum, behavioural results suggest that participants were successfully preparing an action following S1, in both cases of intentional and instructed trials. As it was reported by Fleming et al. (2009), there was no direct behavioural evidence to suggest that intentional trials were more susceptible to task-irrelevant stimuli than instructed trials.

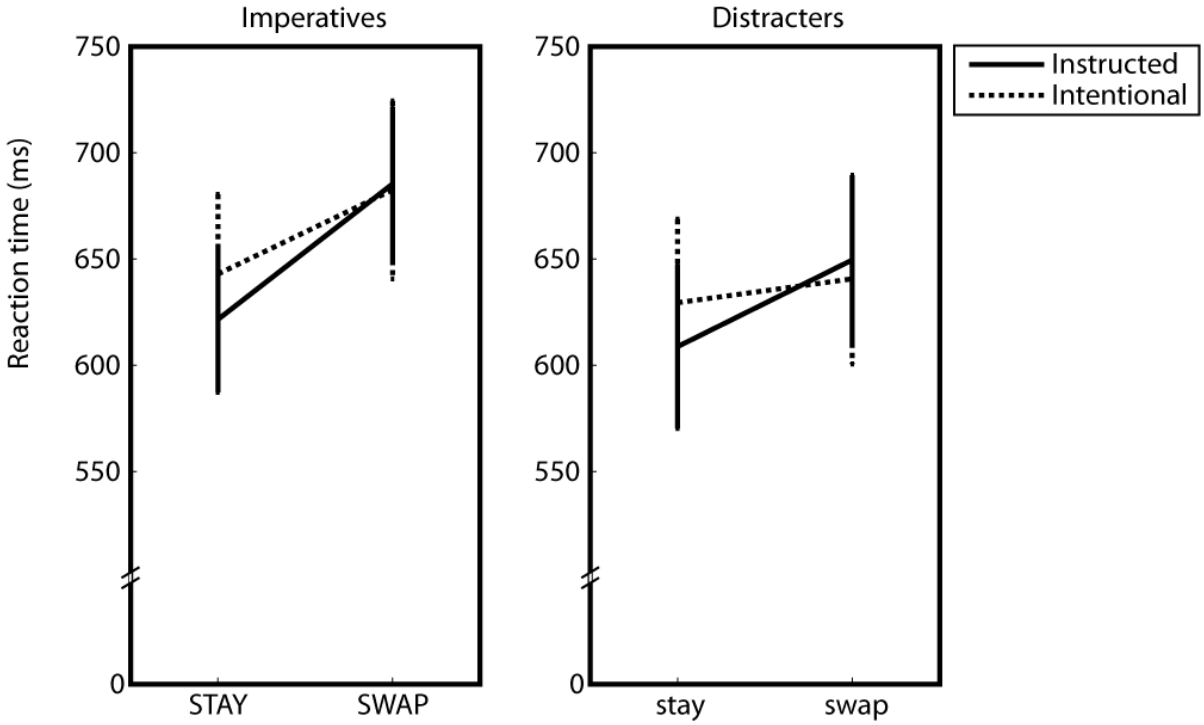


Figure 5.4 Mean RTs for **A** task-relevant and **B** task-irrelevant S2 trials. There was a significant *swap* RT cost, present in both instructed and intentional trials. Error bars show standard error.

EEG results

As an initial approach to identifying the timing of the ERPs, trials were time locked to S1 and examined in their full length. Figure 5.5 shows the grand average ERP for all conditions collapsed, measured from electrode Pz. The grand average trace shows two clear P300 components, peaking at around 350-400 ms after stimulus onset. Topographical distributions of ERPs were consequently obtained at 400 ms after S1 and S2 onset, respectively. The P300 components have a clear topographical distribution centred on electrode Pz. Thus, all subsequent analyses were done on the EEG signal recorded from Pz.

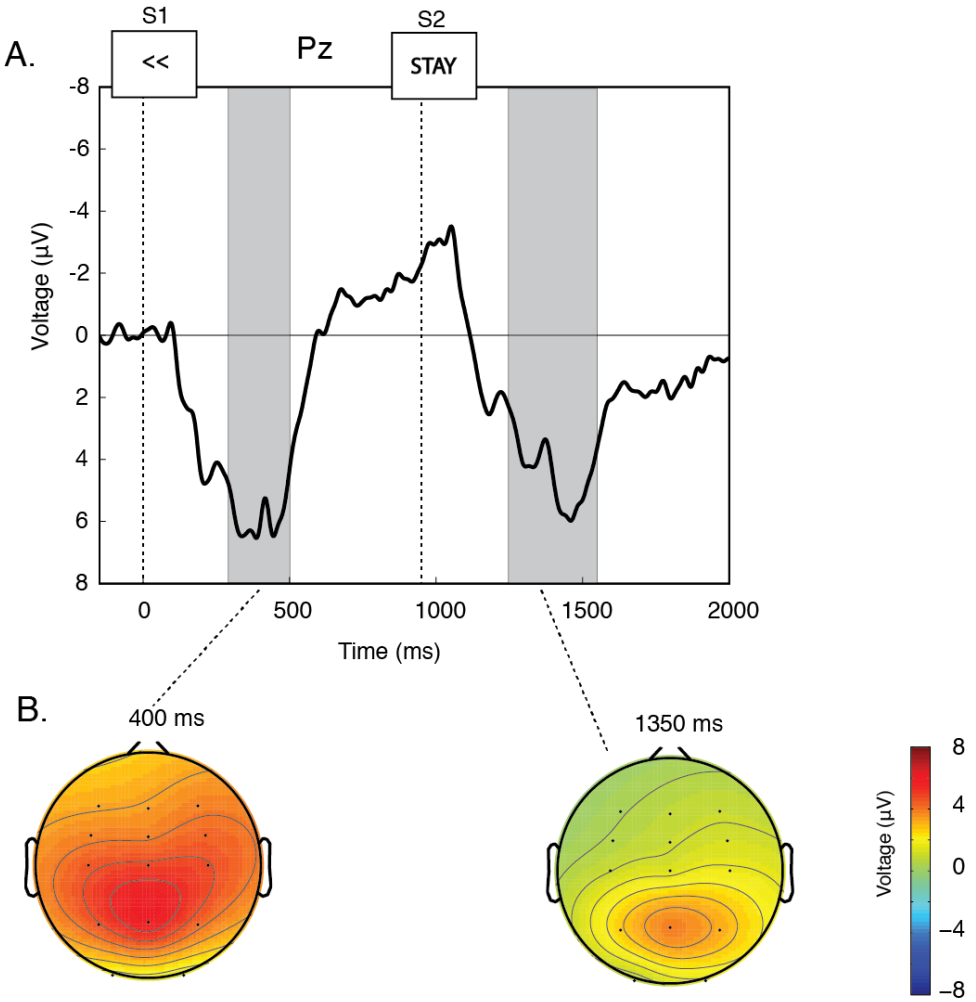


Figure 5.5 A Grand average ERP in electrode Pz for all 8 S-S2 conditions together, time-locked to S1. Vertical dotted lines indicate the time of onset of S1 (0 ms) and S2 (950 ms) respectively. Insets show an example S1 and S2. **B** Topographical plots at 400 ms after visual signal onset (neural response to S1: 400 ms; neural response to S2: 1350 ms). The scalp distributions show that both P300 potentials were clearly centred on Pz.

Neural response to S1

At the time of S1, trials could only be classified as either instructed or intentional, as S2 information had not yet been provided. After blink rejection, an average (± SD) of 200 ± 44 trials remained in the instructed conditions, and 215 ± 42 trials remained in the intentional conditions. There were significant differences between the numbers of trials for instructed and intentional conditions (t_{11}=-3.37, p=0.006). Differences in trial numbers may lead to differences in estimated ERP amplitudes (Luck, 2005). However, increased numbers of trials will typically lead to increased ERP amplitudes. Here, although more trials were classified as intentional, this condition

showed smaller ERP amplitudes (see below). Therefore the differences in trial numbers would have decreased any potential significant differences the ERP amplitude between conditions and cannot directly account for these results.

The mean S1 P300 amplitudes were compared between instructed and intentional conditions (see figure 5.6). On the basis of the grand average of all conditions, the mean S1 P300 amplitudes were measured in the time window between 300 and 500 ms after S1 onset, in electrode Pz. S1 P300 amplitudes were significantly larger for instructed conditions as compared to instructed conditions (t_{11}=3.24, p=0.008).

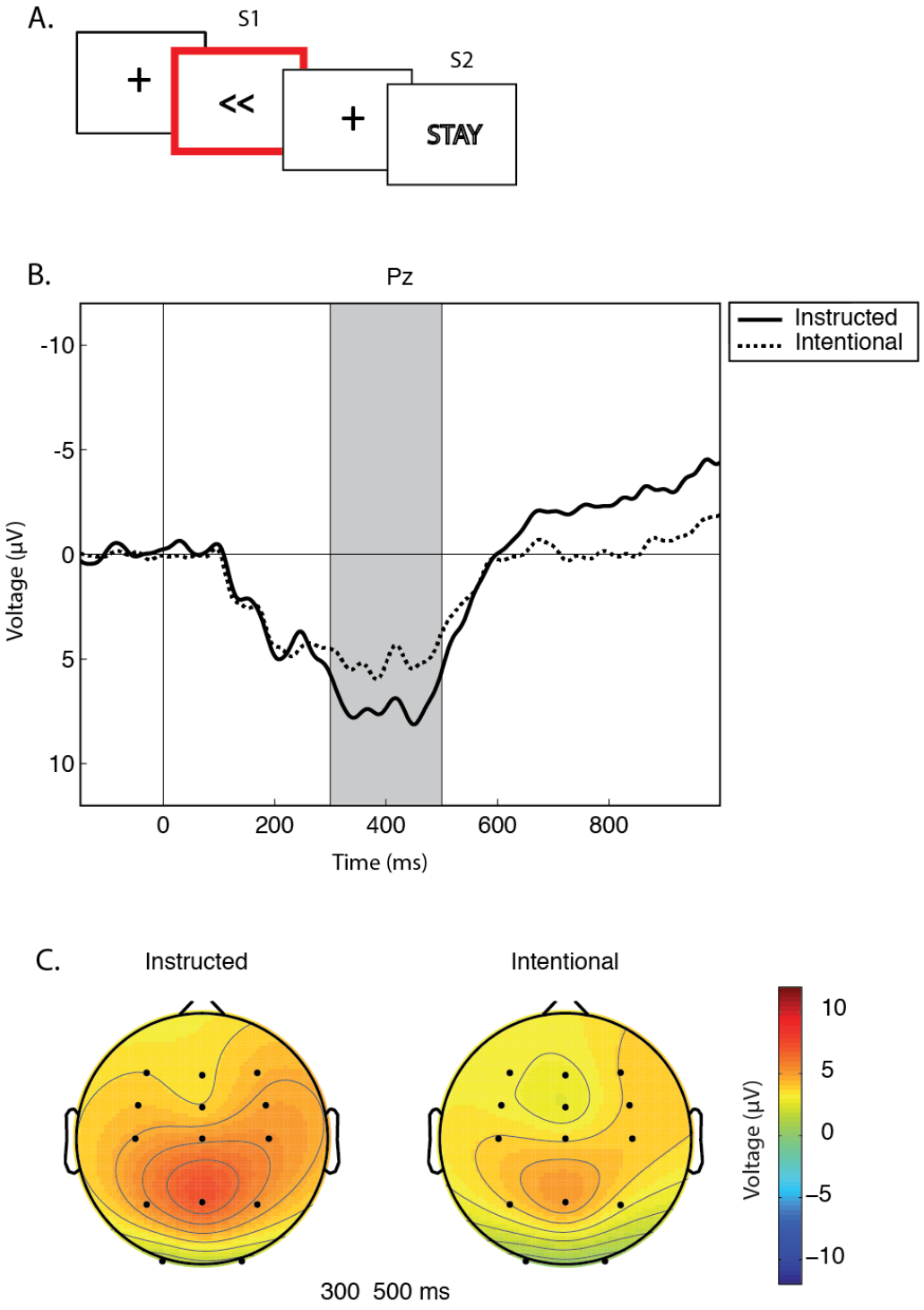


Figure 5.6 A. Neural response to S1.At the time of S1 trials can only be classified as instructed (<< or >>) or intentional (<>). **B.** ERPs time locked to S1. **C.** Topographical distribution of the S1 P300 in the time window between 300 and 500 ms after S1 onset. P300 potentials are clearly centred around Pz

These results agree with those reported by Fleming et al (Fleming et al., 2009), where instructed trials also elicited a stronger P300 amplitude, albeit the maximal difference was slightly later in their original report, at 520-540 ms.

Neural response to S2

The EEG signal was time locked at S2 onset, and trials with blink artefacts were removed. Because 8 conditions could be defined, there were less trials per condition in S2 than in the case of time locking to S1. The data from one additional participant had to be discarded because there were less than 10 trials in some conditions, yielding a total of 11 participants for the ERP analyses at the time of S2. After blink rejection, the mean (± SD) numbers of trials per condition were as indicated in table 5.1

Table 5.1 Mean numbers of trials (± SD) in each condition, for ERPs time locked to S2

Condition	Number of trials
Instructed task-relevant *swap*	49 ± 18
Instructed task-relevant *stay*	51 ± 7
Instructed task-irrelevant *swap*	50 ± 6
Instructed task-irrelevant *stay*	55 ± 8
Intentional task-relevant *swap*	54 ± 5
Intentional task-relevant *stay*	56 ± 7
Intentional task-irrelevant *swap*	55 ± 7
Intentional task-irrelevant *stay*	53 ± 4

A three-way ANOVA with the factors of source, condition and outcome was performed on the mean numbers of trials per condition. There was a small but significant main effect of source ($F_{1,11}$=6.49, p=0.029). No other significant differences were found between the numbers of trials for each condition. The intentional conditions included more trials than the instructed conditions. Importantly however, as in the case of S1, the S2 P300 amplitudes were larger for

instructed trials than for intentional trials (see below). Therefore, differences in trials numbers between instructed and intentional conditions cannot directly account for the pattern of ERP results obtained.

S2 was a more complex stimulus than S1, as it contained both semantic information (*stay*/*swap* instructions) and non-semantic information (uppercase/lowercase, indicating the relevance of S2). This may explain the latency and profile differences between S1 and S2 P300s. On the basis of the grand average, the time window of 300-600 ms was considered for the analysis of the S2 P300 amplitude.

First, to evaluate whether the result reported by Fleming et al was replicated here, only task-relevant S2 conditions were analyzed. These conditions were effectively identical to the original experimental design in Fleming *et al* (2009). Mean P300 amplitudes were compared between *swap* and *stay* task-relevant trials for conditions corresponding to either source of decision. A priori t-tests revealed a significant difference between *swap* and *stay* trials in the instructed conditions (t_{10}= 2.49, p=0.032). Conversely, P300 amplitudes for *swap* and *stay* trials in the intentional conditions did not differ significantly (t_{10}= 1.03, p=0.335). These results are therefore in line with what was reported by Fleming *et al* (2009).

The main experimental question of whether instructed and intentional conditions differ in terms of their relative processing of task-irrelevant stimuli was addressed next. The EEG signal was time-locked at the time of S2 onset and the amplitude of the S2 P300 was measured for each condition. 8 different conditions were considered, resulting from the three-way factorial design. The ERP traces time locked to S2 for each condition are shown in figure 5.7, and the topographical scalp distributions are shown in figure 5.8.

A three-way ANOVA on the mean window amplitudes revealed a main effect of source ($F_{1,10}$=31.34, p<0.001), which agrees with what had been reported by Fleming et al. There was no main effect of relevance ($F_{1,10}$=1.80, p=0.209) and no main effect of outcome ($F_{1,10}$=1.12, p=0.314). These results contrast with those obtained from the behavioural data (see above). An analysis of the RTs for the S1-S2 conditions had revealed significant main effects of both outcome and relevance. These results, and in particular the main effect of outcome, had been taken to suggest that there was

some automatic processing of the *swap* stimuli, both in the instructed and intentional conditions. These results, however, were not reflected in the P300 amplitudes.

There was no significant source x outcome interaction ($F_{1,10}$=1.98, p=0.190), but a significant relevance x outcome interaction ($F_{1,10}$=5.57, p=0.040). There was no significant three-way interaction ($F_{1,10}$=0.22, p=0.647).

There was a significant source x relevance interaction ($F_{1,10}$=5.53, p=0.041). This interaction was further explored by collapsing across the two possible outcomes. Post hoc t-tests revealed that whereas there was a trend for significant differences between the instructed conditions (t_{10}=1.92, p=0.083), the intentional conditions were far from reaching significance levels (t_{10}=0.58; p=0.573).

Table 5.2 offers the results of the 2x2x2 ANOVAs performed on either the mean RTs or the mean P300 window amplitudes. Behavioural and EEG analyses revealed strikingly different results, suggesting that they reflect different underlying processes.

Table 5.2 Results of the 2x2x2 ANOVA on the behavioural and electrophysiological data time locked at S2.

	RT		**P300 mean amplitude**	
Effect	**$F_{1,11}$**	**p**	**$F_{1,11}$**	**p**
Source	0.341	0.571	31.34	<0.001*
Relevance	5.85	0.034*	3.80	0.209
Outcome	21.48	<0.001*	1.12	0.314
Source x relevance	0.143	0.712	5.53	0.041*
Relevance x outcome	6.04	0.032*	1.98	0.190
Source x outcome	3.16	0.103	5.57	0.040*
Source x relevance x outcome	0.11	0.748	0.22	0.647

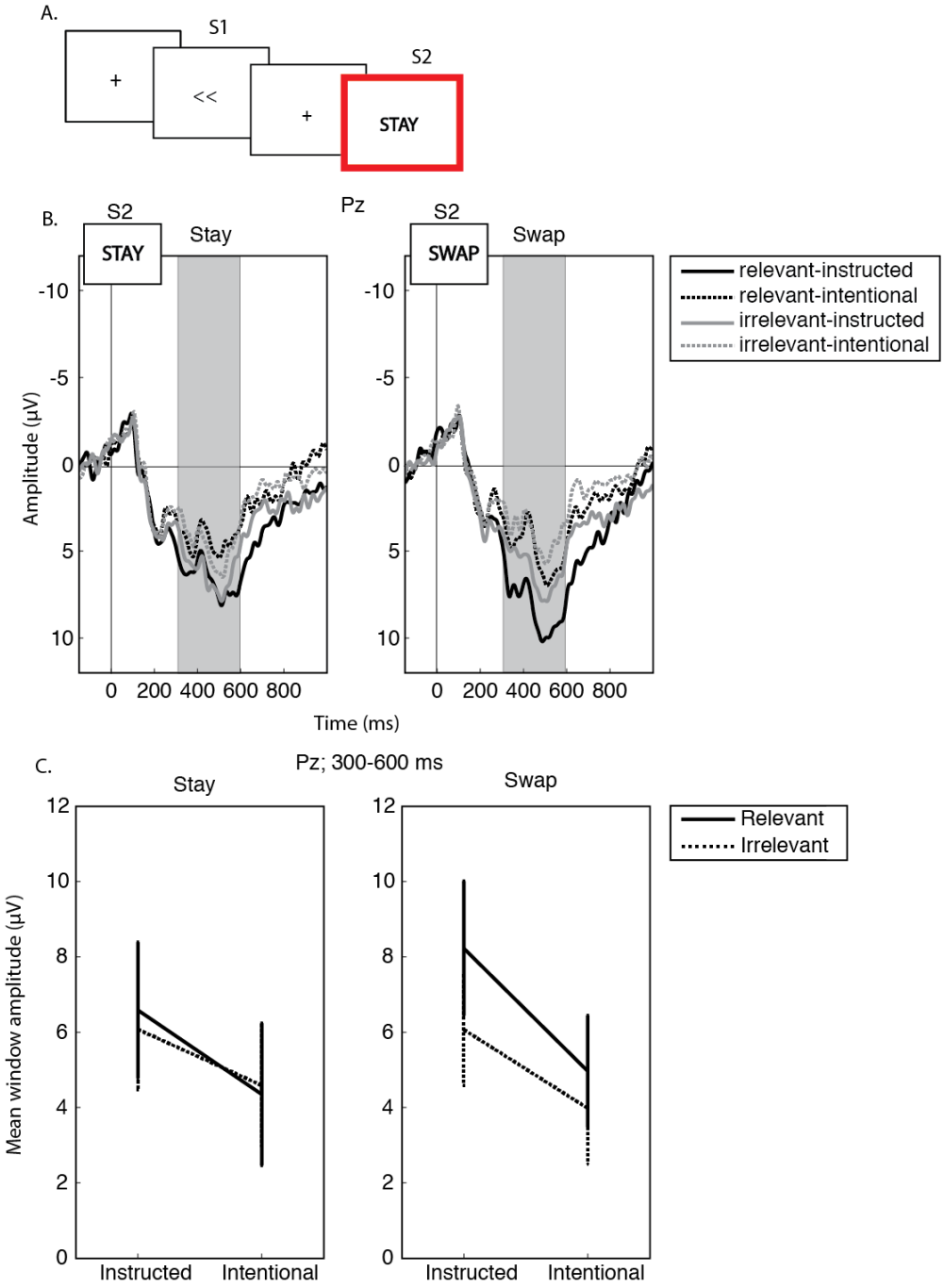


Figure 5.7 A. Neural responses to S2. **B.** ERPs to each of the 8 S1-S2 conditions. **C.** Mean ERP amplitudes were obtained from a time window between 300 and 600 ms after S2 onset from electrode Pz.

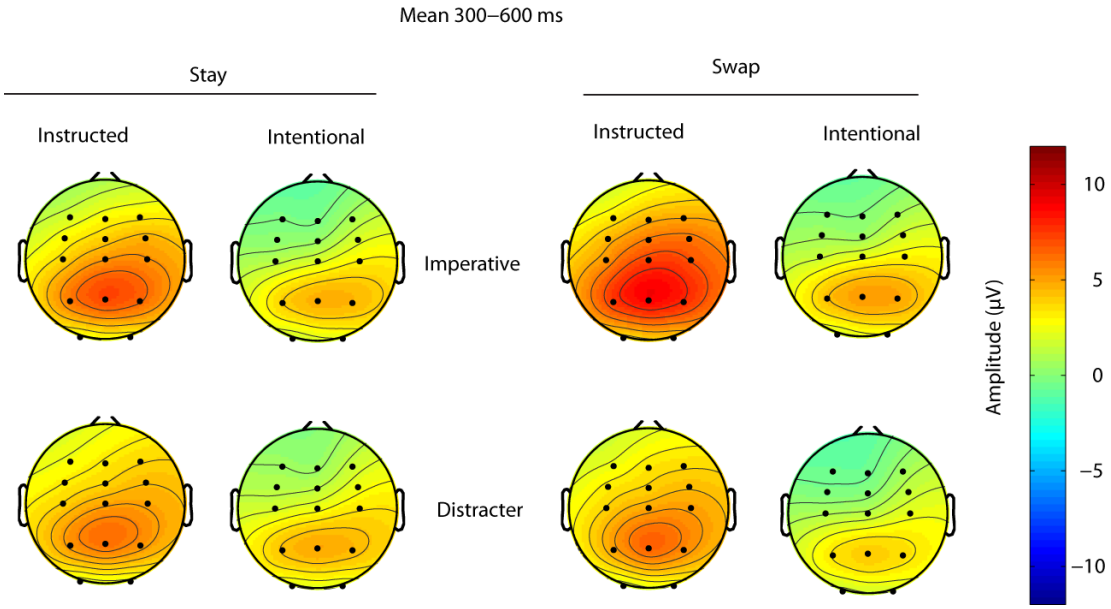


Figure 5.8 Topographical distribution of the neural response to S2, in a time window of 300 to 600 ms after S2 onset. As in the the case of S1 P300, S2 ERPs were centred around Pz.

Evidence for motor preparation following S1

The analysis of RT data suggests that participants were indeed preparing a directional motor response, in both intentional and instructed trials. To seek for neurophysiological evidence for response switching following S2 switch cues, the motor-related amplitude asymmetry (MRAA) was computed (see methods section). Although there was an expected small decrease in μ-band power around the time of S2 cue, this decrease was comparable for contra- and ipsilateral electrodes. MRAA calculations provided no strong evidence for lateralization of action preparation or switching of action plans following task-relevant S2 *swap* cues (see figure 5.9).

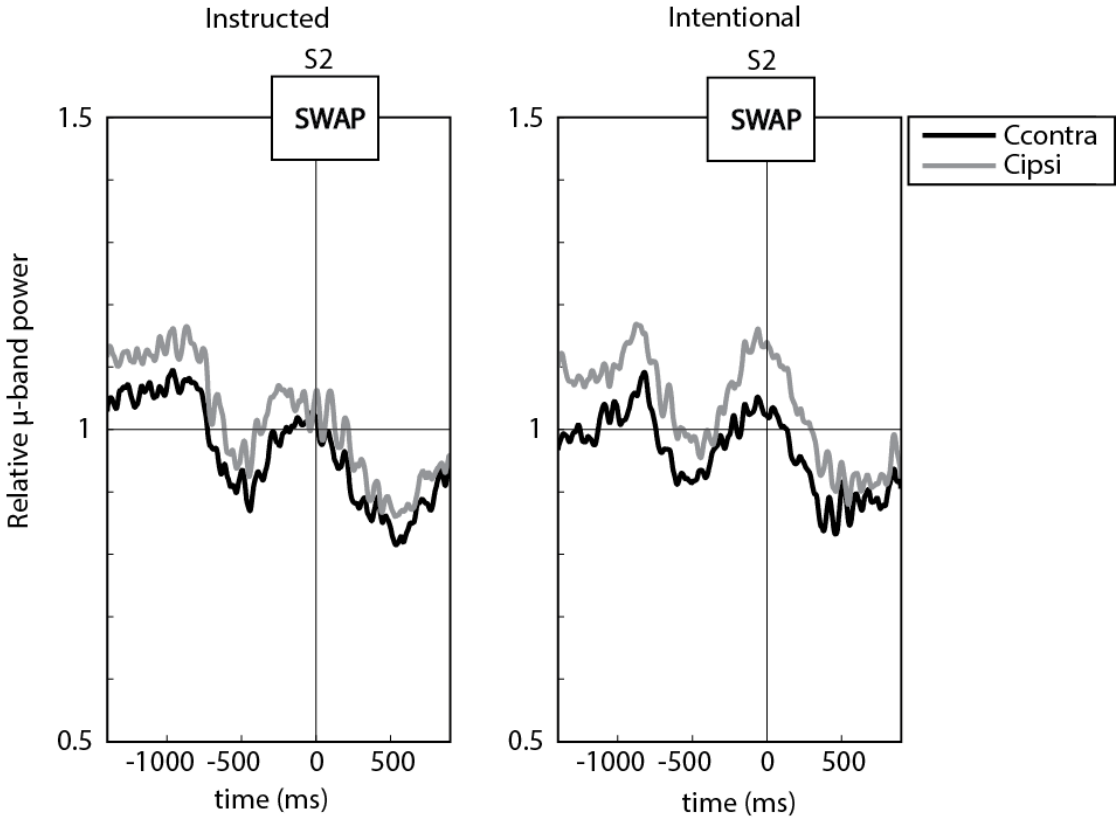


Figure 5.9 No evidence for switching of action plans following S2 cues. Relative μ-band power showed a slight decrease around the time of action. There was no inversion of the polarity of the differences between contra- and ipsilateral C electrodes following an S2 *swap* cue, as expected if participants switched action plans.

Controlling for incorrect intentional trials

In this paradigm, the intentional decision following S1 was covert, and was expressed only through the response following S2. Thus, potential incorrect *swap* or *stay* intentional trials could not be detected. Because incorrect trials were rejected from the instructed conditions but not from the intentional conditions, the analyses reported above introduced an asymmetry in the analysis of instructed and intentional conditions. To account for this potential confound, the same ERP analysis was done for S2, without rejecting the incorrect instructed trials. An analysis of S2 ERP amplitudes revealed virtually identical results as the ones reported above.

As in the previous analysis that did not include the ~8% incorrect instructed trials, the ANOVA of the mean ERP amplitudes following S2 revealed a main effect of source ($F_{1,11}$=31.06, p=0.002), no main effect of relevance ($F_{1,11}$=2.92, p=0.115) and no main effect of outcome ($F_{1,11}$=1.46, p=0.251). The two- and three-way

interactions also showed the same effects as in the previous analysis. There was a significant source x relevance interaction ($F_{1,11}$=6.72, p=0.025), no significant source x outcome interaction ($F_{1,11}$=3.46, p=0.090), a significant relevance x outcome interaction ($F_{1,11}$=6.58, p=0.026), and no significant three-way interaction ($F_{1,11}$<0.001, p>0.999). Therefore, including incorrect trials in the analysis of instructed trials did not change the results qualitatively.

5. 4 Discussion

Results described in previous chapters of this thesis suggest that intentional decisions have weaker neural representations than their instructed equivalents, especially during periods of action preparation. One previous experiment had suggested that intentional action plans are more flexible than instructed action plans (Fleming et al., 2009). The study presented here aimed at further probing the period of action preparation by extending that result. It addressed the relative susceptibility of instructed and intentional action plans to external task-irrelevant stimuli. The relative processing of task-irrelevant stimuli was measured for situations of action preparation of both instructed and intentional decisions. Results revealed that when instructed action preparation is underway, the differences in the processing of task-relevant and task-irrelevant stimuli is stronger than that in situations in which intentional action preparation is underway. These results are consistent with a scenario in which the distinction between task-relevant and task-irrelevant stimuli was efficiently made when instructed actions plans were underway, but not as efficiently when intentional actions are prepared.

Neural responses to S1

First, the amplitude of the P300 elicited by S1 was compared across conditions. S1 only specified the source of action (instructed/intentional), so only these two conditions could be compared at the time of S1. In agreement with what had been

reported before, instructed S1 stimuli elicited stronger P300 amplitudes than intentional S1 stimuli (Fleming et al., 2009; Waszak et al., 2005).

As Fleming *et al* (2009) have argued, at least two possible explanations can account for this result. First, differences in S1 P300 amplitude may be simply related to differences in S1 cue frequency. To balance the number of intentional and instructed trials, the experimental design necessarily included twice as many instances of the intentional S1 cue (<>) than those with either of the instructed S1 cues (<< or >>). P300 amplitude increases with increased stimulus novelty, reflecting adaptation of the neural response (Goldstein, Spencer, & Donchin, 2002; He, Lian, Spencer, Dien, & Donchin, 2001). Therefore, less exposure to either one of the instructed S1 cues could have led to an increased S1-P300. However, novelty effects on P300 amplitudes are normally observed in "oddball" paradigms, where frequency ratios of frequent and oddball stimuli are typically around 9:1, much larger than the 2:1 ratio of the paradigm used here. An alternative explanation, favoured by Fleming *et al* is that the observed amplitude differences are indeed related to differences between instructed and intentional choice processes.

As an additional alternative explanation, increased P300 amplitudes may reflect more bits of information extracted from the S1 cue, as it has been suggested to be the case for other visual stimuli (Duncan-Johnson & Donchin, 1982; Gratton et al., 1990). Because instructed S1 cues specify the correct key press direction, they convey more information than intentional S1 cues, which leave the direction dimension underspecified. More information conveyed by a cue may translate into larger ERP amplitudes, as a result of more neural processing.

Neural responses to S2

The amplitude of the ERPs elicited by S2 was compared across conditions. Here, all 8 S1-S2 conditions could be resolved. Three factors were manipulated independently. The S1 cue determined the source of decision (instructed/intentional), and the S2 cue manipulated both the relevance of the S2 cue (task-relevant/task-irrelevant) and the action outcome (*stay*/*swap*).

In a design involving only task-relevant S2 cues, Fleming *et al* had reported an interaction effect between source of decision and S2 instruction outcome, with greatest S2-P300 amplitudes in instructed-*swap* trials, and no differences between the *stay* and *swap* trials in the intentional conditions. Here, these findings were replicated, as revealed by *a priori* testing. The authors interpreted these results as increased neural effort required for the restructuring of instructed action plans as compared with intentional action plans.

In the present study, the main hypothesis was that intentional decisions would show increased processing of task-irrelevant stimuli as compared to instructed decisions. In order to test this hypothesis, the S2 P300 amplitudes were compared across the 8 conditions. The critical finding was that the S2 P300 amplitude showed a statistically significant interaction between the factors of relevance and source of decision. The differences between the task-relevant and task-irrelevant stimuli in the instructed condition were stronger than in the intentional conditions. This suggests that in the intentional conditions, the distinction between task-relevant and task-irrelevant stimuli was made less efficiently than in instructed conditions.

These results remained significant even after controlling for a potential confound due to unidentified incorrect intentional trials.

Processes affecting P300 amplitude

The amplitude of the P300 component, measured here, has been related to several different cognitive processes. Under one influential view, P300 amplitude relates to "context updating" (Donchin & Coles, 1988). Within this framework, P300 amplitudes correlate with the amount of neural "effort" necessary to update action plans. Larger P300 amplitudes may be indirect measures of the distance between the two neural representations of action plans, because more neural activity is required to change from one to the other.

An additional alternative account should be considered here. As it was argued in the case of S1, differences in S2 P300 amplitude could be related to the amount of information extracted from a stimulus (Duncan-Johnson & Donchin, 1982; Gratton et al., 1990). In this task, the processing of S2 could be done at two different levels.

First, S2 stimuli had to be identified as task-relevant or task-irrelevant, on the basis of their surface form (i.e., the letter case). Also, only for those trials with task-relevant stimuli, the semantic content of S2 had to be evaluated to interpret the S2 instruction and produce the adequate behaviour. The results presented here are also consistent with this account. Importantly however, to the extent that there was true action preparation in the intentional conditions (as the behavioural data suggests) intentional and instructed conditions should not have differed in terms of the amount of information extracted from the S2 cue. Therefore, the information extraction account alone cannot easily explain the results reported here.

In this experiment, reaction times and P300 amplitudes showed strikingly different sensitivities, suggesting that they reflect different processes. In go/nogo tasks, N2/P3 complexes have been typically related to response inhibition (see introduction, section 1.3.5). However, the two subcomponents are thought to serve different functions. The N2 component amplitude is related to nogo signal monitoring (or, more generally, conflict monitoring) monitoring. The P3 component amplitude, on the other hand, has been related to actual response inhibition (see section 1.3.5, Bekker, Kenemans, Hoeksma, Talsma, & Verbaten, 2005; Dimoska, Johnstone, & Barry, 2006; Smith, Johnstone, & Barry, 2008). Speculatively, then, one possible interpretation of the discrepancies between the RT and ERP data might suggest that RT modulations in task-irrelevant *swap* trials are due to automatic conflict monitoring. These increased RTs would therefore correlate with N2 amplitudes, but not with the P300 amplitudes, measured here.

Evidence for moment preparation and switching of action plans

In the intentional conditions, participants could freely choose which hand they would use to execute their response. The choice following S1 was intentional, and it was not revealed to the experimenter apart from through the behavioural outcome. Therefore incorrect trials could not be identified. Thus it is in principle possible that participants were not preparing action in intentional trials, and simply pressed any key they wished following S2. Under this account, P300 amplitude may not have

differed between task-relevant and task-irrelevant stimuli because after an intentional S1 participants could have simply ignored all subsequent visual cues, effectively eliminating any differences between the S2 conditions.

Perhaps the most convincing evidence against this possibility would come from neurophysiological measures of lateralized action preparation and switching behaviour following S2 *swap* cues. Here, the typical measure of lateralization of response preparation, namely the LRP, could not be calculated. The processing of the visual cue S2 occurred at variable times relative to the time of key press, confounding the LRP amplitude with the processing of the S2 cues. Instead, MRAA was calculated as a measure of lateralization (de Jong, Gladwin, & Hart, 2006; Gladwin et al., 2006). This measure provided no evidence for lateralized action preparation in either instructed or intentional conditions. Thus, on the basis of the neurophysiological data alone, no conclusions can be drawn about the nature of action preparation or successful changing of action plans.

Still, three pieces of behavioural evidence suggest that action preparation did indeed follow the S1 cue in the intentional conditions. First, as in the case of Fleming et al, this experiment included "short" trials to encourage action preparation. In these trials participants had to execute their prepared action. Mean RTs in short trials served as an indirect measure of action preparation following S1 in S1-S2 trials. Results revealed significant differences between baseline trials (where no directional action preparation could have occurred) and short trials (where action preparation was encouraged). This was also (and especially) the case for intentional trials, suggesting that participants were not simply ignoring the S1 cues in the intentional conditions.

Further, behavioural evidence also suggests that participants did indeed follow the S2 *stay*/*swap* instructions. When RTs were compared across all 8 conditions, a significant effect of outcome emerged, and although the factor of outcome (*stay*/*swap*) did interact with relevance (instructed/intentional), follow up comparisons revealed that *swap* trials showed significantly longer RTs than *stay* trials in both instructed and intentional trials. It must be noted however that although task-relevant trials were in general associated with longer RTs, there was no evidence for a significant interaction effect between the factors of relevance and outcome. Intriguingly, this suggests that the RTs were as least in part modulated by

the processing of the S2 *swap* cues in general, regardless of the actual need to switch action plans.

Importantly, the amplitudes of the P300 elicited by S2 presented a somewhat different pattern of results. In the case of S2 P300, there was a significant main effect of outcome that did *not* interact with the S2 cue relevance. These opposing patterns of results for RTs and S2 P300 amplitudes suggest that the two measures reflect two different processes. Increased RTs may reflect increased or deeper cognitive processing of a stimulus whereas the S2 P300 may reflect neural effort required for a switch in action plans, as Fleming *et al* had originally suggested.

Limitations of this study

The most conspicuous limitation of this experimental design is that its factorial nature depended heavily on the participants' behaviour. Instructed S1 cues unambiguously signalled which action (right- or left-hand key press) participants had to prepare. However, in the intentional conditions the decision about which action to select was left to the participants' choice. Thus, it could not be evaluated solely by behavioural measures whether participants were in fact following the instructions indicated by the S2 cue.

A further limitation is related to the choice of word S2 stimuli, and the assignment of upper-and lower-case stimuli to the instructed and intentional conditions. The purpose of the S2 cues was twofold. S2 were designed to be easy to identify, based on their surface form and not their semantic content, and at the same time difficult to ignore, due to automatic semantic processing. However, it is possible that implicit associations between e.g. uppercase and task-relevant conditions occurred, introducing potential confounds in the data. Because these conditions were counterbalanced between and within each experimental session, any potential systematic differences would have been counterbalanced but would have also decreased any potential differences between conditions.

Implications of this study

Automatic processing of external signals has been studied in fields such as advertisement research. For example, Treleaven-Hassard *et al* (Treleaven-Hassard et al., 2010) have recently measured P3a latency as a means to measure the effect of different advertising techniques on the automatic processing of the advertised brands. The authors measured the neural response to visual stimuli to reveal differences in the underlying mental states (like/dislike of a product). If it is true that intentional decisions are weaker in suppressing the influence of irrelevant stimuli, an interesting corollary of this study may be that freely-chosen alternatives are potentially relatively easy to influence, and highly susceptible to external suggestion.

5. 5 Conclusion

The examination of the period of action preparation is crucial in testing the hypothesis that instructed and intentional decisions differ in their relative strengths. One possible approach to measuring the strength of action preparation is to do it indirectly, measuring in turn the relative processing of task-irrelevant stimuli in periods of action preparation. The results presented here provide evidence compatible with the notion that intentional decisions lead to less efficient suppression of task-irrelevant stimuli as compared to instructed decisions, lending support to the hypothesis that intentional decisions have weaker neural codes than instructed decisions.

Chapter 6 Resistance of the internal representation of response alternatives

"There may be outward impediments even whilst deliberating, as a man deliberates whether he shall play at tennis, and at the same time the door of the tennis court is fast locked against him. And after a man has ceased to deliberate, there may be no outward impediments, as when a man resolves not to play at tennis because he finds himself ill-disposed, or because he will not hazard his money. So the same person, at the same time, should be free and not free, not free and free" (Bramhall, 1655)

Accounts of intentional action often refer to unchosen response alternatives: "could have done otherwise". Chapter 6 focussed on the internal representation of these response alternatives for intentional action. Classically, reaction times increase with the number of alternatives. In this behavioural experiment, when possible responses were removed from the stimulus set, reaction times correlated better with the original number of possible responses than with the updated number of responses after removal. Internal representations of response alternatives lag behind actual alternatives. Effectively unavailable alternatives may still be "internally" contemplated. "Could have done otherwise" may be linked to a perseverative tendency to maintain in working memory choices, which we are in fact no longer free to make.

6. 1 Introduction

Intentional choices may not be deterred by instructed inhibition

When a person chooses between alternative actions, what determines how widely her choices range? What is the domain over which action selection mechanisms select? How is this domain dynamically updated, as new action options appear and disappear? A common example may illustrate this question. After browsing the menu, a customer in a busy restaurant asks for the dish of the day, only to be later informed that her choice is no longer available. The disappointed customer may then have to re-choose from a now reduced set of available dinner alternatives. Further, the costumer's potential disappointment might not have occurred if the chosen dish had never been included in the menu. Arguably, the reason why the *late* external restriction on the menu led to disappointment and the *early* restriction did not is the contemplation of the possibility. In the former case, the "dish of the day" response alternative had been internally represented in the costumer's brain. Conversely, in the latter case, this option had never been represented, and it therefore does not carry any affective value. This everyday example illustrates two processes of interest in the scope of this thesis. These are, on the one hand, the internal representation of intentional response alternatives, and on the other hand, the interaction between intentional selection and instructed (external) restrictions on the intentional space of response alternatives.

Intentional decisions rely on internal representations of available alternatives

Chapters 2 and 3 in this thesis suggested that the neural representations of the alternative and unselected outcomes may remain partially activated even after an intentional decision for a particular course of action has been made. This may not happen to the same extent for instructed decisions, where the external environment clearly precludes the alternative course of action, leading to a more robust inhibition of any neural representations of "incorrect" response alternatives. In light of this interpretation, the study of the internal representations of response alternatives in

cases of intentional actions becomes a matter of critical interest in the scope of this thesis.

To further examine this question, this study evaluated the persistence of the internal representation of response alternatives in cases of intentional action. The interaction between the intentional selection of action and late instructed restrictions on the available set of alternatives was investigated. In particular, this study asked whether the internal representations of response alternatives could flexibly follow external changes in the environment that restrict the effectively available alternatives.

Behavioural manifestations of response set size

The sizes of the sets of available response alternatives can be evaluated through their behavioural effects. Hick (Hick, 1952) examined the relationship between participants' reaction times (RTs) to press a button and the number of potential response alternatives. He tested participants in an apparatus that could automatically turn on one out of 10 possible lights. He asked participants to press a key if they saw the corresponding light turn on. Critically, in each experimental block, the number of possible response keys varied. Participants were instructed to rest their fingers on a given subset of the maximum ten possible keys. In that way, the *sizes* of the response set varied across blocks, from 1 to 10. The number of potential response alternatives constitutes the response set size. Hick reported a monotonically increasing relationship of RT with the response set size, which could be optimally fit by a binary logarithmic function, now widely known as Hick's law.

Hick's law was drawn from measurements of instructed action RTs. Intentional action RTs have also been found to increase with increasing response set sizes, although a strict logarithmic relationship has not been tested (e.g., Kühn, Gevers, & Brass, 2009; Lau, Rogers, Ramnani, & Passingham, 2004; van Eimeren et al., 2006; Zhang, Hughes, & Rowe, 2012).

Internal vs. external response sets

In his experiments, Hick explicitly informed his participants of the external set of available responses, so that they knew how many possible responses there were. However, and importantly, this external response set cannot influence behaviour *per se*, but only through its internal representation. Therefore, a crucial distinction must be made, between *external* and *internal* response sets. The former refers to the response alternatives that are effectively available in the external environment. The latter refers to the internal representations of these response alternatives.

Hick could safely assume that the internal response set matched the external response set in his experiments. Importantly however, the external and internal response sets are not necessarily equivalent. For example, a person may internally consider response alternatives that are in fact unavailable in the external environment. Such would be the case of a completely delusional misalignment between the internal and external response sets, in a person considering unrealistic action alternatives. A second, more realistic example would be the everyday scenario described in the introduction of this chapter. In that case, the internal representation of the possible alternatives initially contained one more item (the "dish of the day") than the external response set. Alternatively, internal representations may represent a restricted subset of the externally available alternatives, in cases of vast external response sets. For instance, in a word completion task (e.g., Jacoby, Toth, & Yonelinas, 1993), an efficient strategy may be to select a response from a restricted set of words, instead of searching the full lexicon.

Do internal response sets follow changes in external response sets?

This study aimed at probing the relationship between the external and internal response sets. In particular do internal response sets closely follow sudden changes in the external response set, or are they instead resistant to change? An experimental task was designed based on the restaurant scenario described above. Participants were asked to intentionally select a response from an initially available set of responses. Once an intentional decision had been made, but before the response was executed, the external response set was suddenly reduced. In this way the new and

updated external response set did no longer include the initially selected response. Using RTs as a proxy for the internal response set size, it could be addressed whether the internal response set had been rapidly updated to match the new external response set, or whether the internal response set lagged behind the external changes (see figure 6.1). Two scenarios are possible. In the first place, the internal representation of the response alternatives may represent a perfect and instant match with the external response set. The internal response set may show virtually no delay between the external change and the internal updating. Alternatively, the internal representation of the response alternatives may present some sluggishness in updating. Some delay may be observed between the time of changes in the external environment and the updating of the internal response set.

The hypothesis was as follows. If initially selected response alternatives were effectively removed from the internal response set once they become unavailable, RTs will increase as a function of the final response set size (and not the initial response set size). Conversely, if the neural representation of the selected response alternative is maintained in the internal response set despite it no longer being available in the external response set, then RTs would increase as a function of the initial response set size.

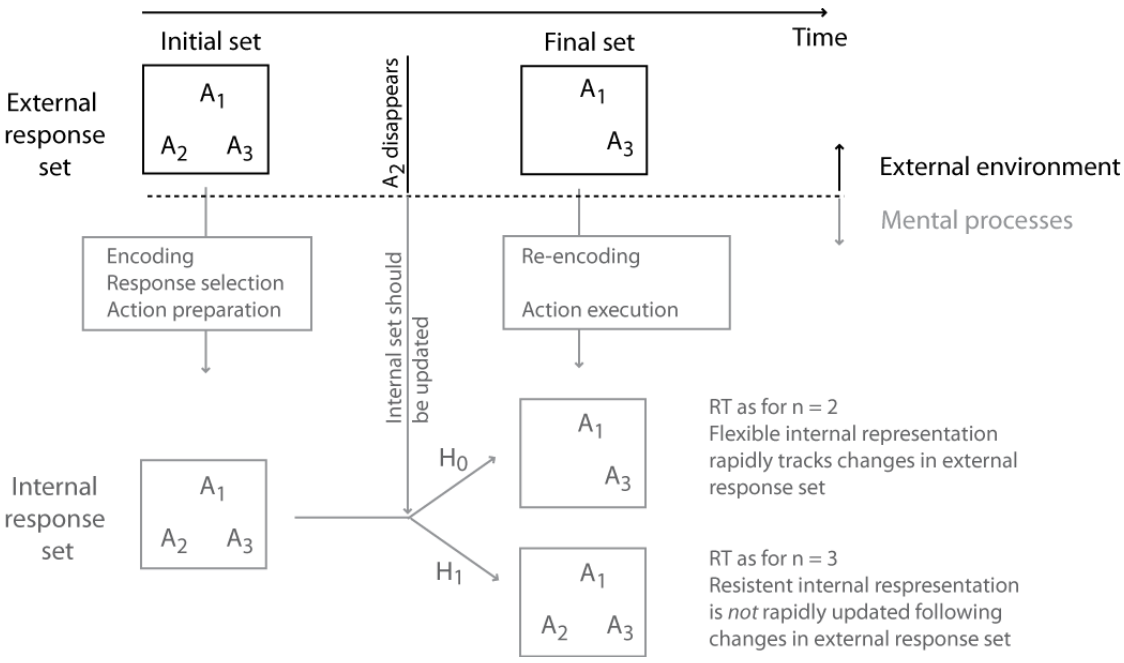


Figure 6.1 Rationale and hypothesis of the study. An initial external response set was presented, containing the response alternatives A_1, A_2 and A_3. Participants were asked to intentionally select and hold a response from this initial response set. In the initial conditions, the internal response set would reflect the external response set. After a sudden change in the external environment, the internal and external response sets would not necessarily match. Two scenarios were possible. The internal response set may be flexible (**A**), and can update its contents to respond rapidly to external changes. Alternatively (**B**), the internal response set may be resilient to change, and the internal response set may lag behind changes in the external environment. The external response set was directly manipulated; whilst the internal response set was measured through its behavioural manifestations (see text for details). As expected by Hick's law, the observed RTs would correspond to the size of either the initial (n=3) or the final (n=2) response sets (where n is number of alternative actions in the response set).

6. 2 Methods

Participants

Eighteen naïve participants (11 female, mean age ± SD; 24 ± 5 years) took part in the study. All participants had normal or corrected to normal vision. Procedures were approved by the UCL research ethics committee and were in accordance with the principles of the Declaration of Helsinki.

Task

Stimuli were displayed on a CRT monitor with a refresh rate of 60 Hz. Participants sat 60 cm away from the screen. The experiment consisted of 6 blocks of 100 trials and lasted for approximately 50 minutes. Each trial belonged to one of four experimental conditions partly determined by the participant's behaviour (see below), *no change* (34% of the total number of trials), *instructed selection* (20%), and *original selection* and *reselection* (together, 46%).

At the start of each trial, one to four different numbers were displayed on the screen, arranged around a central fixation cross with 2º eccentricity (see figure 6.2). All stimuli were displayed over a black background. This set of numbers was the *initial* response set. Numbers in the initial response set were randomly sampled without repetition from the numbers 1 to 9 excluding the number 5 (see instructed condition below). The numbers in the initial response set were displayed in white for either a short or a long stimulus onset asynchrony (SOA). Short SOAs were periods of 350 ms with a random jitter of a maximum of ± 200 ms. Long SOAs were periods of 1500 ms with random jitters of a maximum of ± 200 ms. Participants were asked to covertly select one of the numbers in the initial response set during the SOA, and to prepare to move a cursor and click on the number using a large trackball mouse (Keytools Ltd, Southampton, UK). During short SOAs participants had relatively less time to select a number from the initial response set than in the long SOAs. Short and long SOAs were randomly assigned to experimental trials.

After the SOA, the fixation cross changed colour, from white to red. This was the go signal that indicated to participants that they could start moving the cursor to reach their selected target number. Crucially, simultaneously with the go signal a subset of the numbers in the initial response set disappeared. The remaining numbers changed colour and turned to green. The number of disappearing numbers ranged from none of them to all of the numbers in the initial response set but one (*0* to *n-1*, where *n* is the initial response set size). Consequently, the number of numbers remaining (the *final* response set) ranged from one to the full initial response set (1 to *n*). The positions and identity of the disappearing numbers were fully randomized.

After the go signal, participants could move the cursor to click on the selected number. If the number they had originally selected from the initial response set during the SOA period was still available in the final response set, they could select it. However, if the originally chosen number was no longer present in the final response set, participants were asked to inhibit their planned response and reselect a different number, from the smaller final response set of available alternatives.

After clicking on the number of choice, participants were asked to report which number they had *originally* chosen, regardless of which number they had clicked on. In this way, trials in which reselection had occurred could be identified on the basis of subjective report. If the reported original choice did not match the clicked number, and if the original choice had disappeared, it could be assumed that reselection had occurred. Otherwise, trials were assumed to be simple selection. The use of numbers as targets sought to minimize the working memory load on both target selection and recall, minimizing in turn the problems and potential biases associated with subjective report. At debriefing, no participant reported difficulties in the report of their original choice.

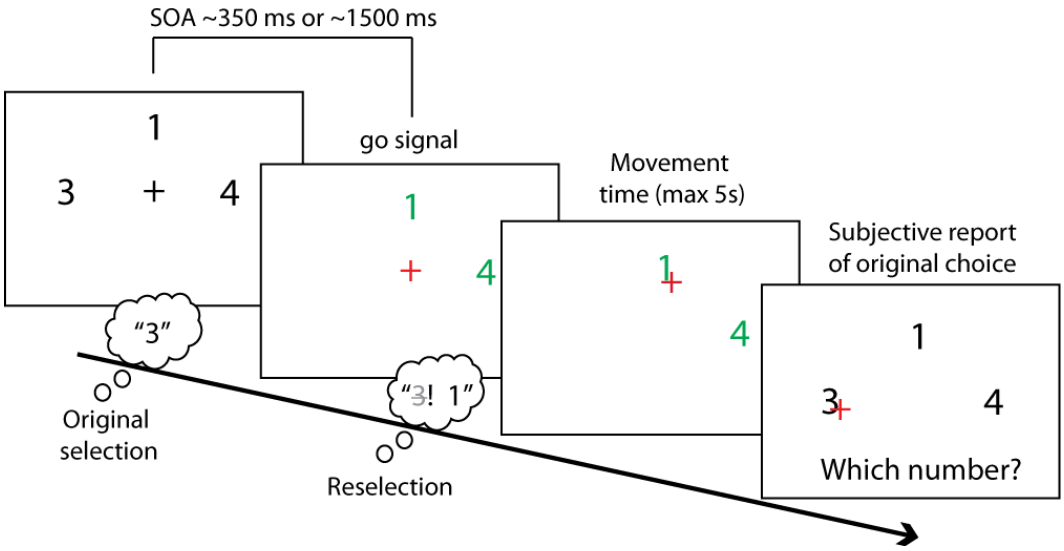


Figure 6.2 Experimental task. **A.** The initial response set of numbers was presented for either a short (350 ± 200 ms) or a long (1500 ± 200 ms) SOA. Participants covertly selected a number during the SOA. A change of fixation colour indicated the go signal. Participants could then move a cursor to click on a number of their choice. At the same time as the go signal onset a subset of the presented numbers could disappear, leaving a final response set with a size between 1 and the original response set size. Participants could click on their originally chosen number if it remained in the final response set. Instead, they would have to reselect a number other than their first choice if it had disappeared from the final response set. Participants then reported their original number choice.

To evaluate the effects of instructed inhibition on the intentional selection from the initially available alternatives, it was crucial that participants did indeed select from the initial response set, rather than simply wait for the appearance of the final response set. Two strategies were used to ensure that participants attended to the initial response set. First, instructed trials were included, which were any trials including a "5" as one of the numbers in the initial response set. In this case, participants were asked to select the number "5" upon seeing the go signal. No numbers were removed from the initial response set in instructed trials, because removing the instructed target in some trials could have decreased the general validity of the instruction. Second, to prevent participants from simply waiting for the final response set, a *no change* condition was included, in which no numbers were removed from the initial response set. Because quick reactions were rewarded, the experimental design discouraged the potential strategy of ignoring the initial response set completely and waiting for the final response set instead.

Further, to encourage action preparation following the initial response set, and therefore the need for inhibition of the prepared response, quick movement times (measured as time to click on the target) were monetarily rewarded. Participants were informed that they would get 0.5 pennies extra for every trial that was quicker than their own average in the preceding block. Participants earned on average (± SD) £ 2.23 ± 0.03.

The large trackball mouse was preferred over a keyboard or a normal mouse for two reasons. First, using a keyboard would have required different effector fingers for each response alternative, possibly introducing systematic RT patterns related to finger dexterity. Positioning the targets on the screen and asking for cursor movements minimized this potential confound. Second, piloting showed that regular mouse or joystick movements were too rapid and familiar to participants. The use of an unfamiliar cursor that required larger movements amplified differences in RTs between conditions.

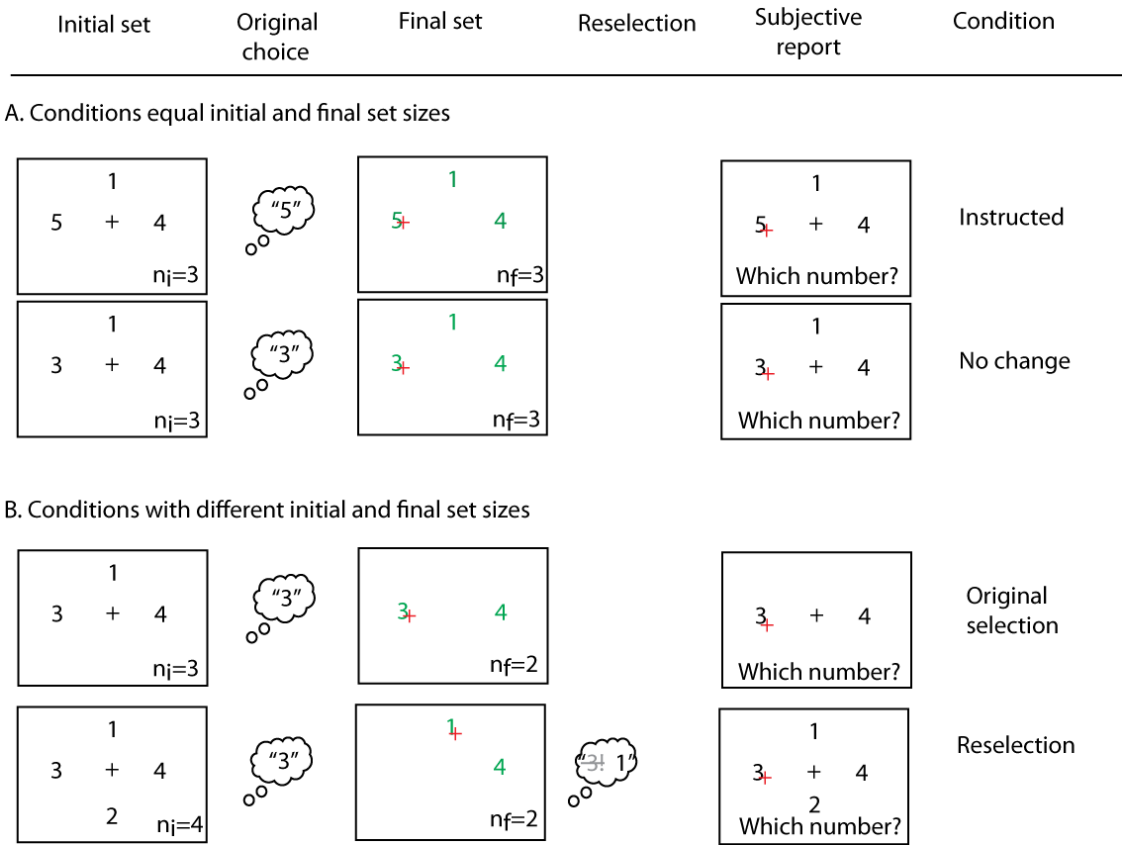


Figure 6.3 Examples of all four experimental conditions**. A.** Conditions with equal initial and final response set sizes. If a number "5" was present in the initial response set, participants were instructed to click on it (*instructed* condition). In the no change condition, the number choice was intentional. **B.** Conditions with non-matching initial and final response set sizes. In the original selection and reselection conditions, some numbers disappeared from the initial response set. A trial belonged to the original selection or the reselection conditions, depending on whether the original intentional choice remained in the final response set, or had been removed from it. A trial was sorted as *original selection* if participants reported that their original choice matched their final choice. Instead, a trial was *reselection* if participants reported having chosen a number that had become unavailable, indicating that they had reselected a number before clicking on their final choice. n_i and n_f indicate the initial and final response set sizes, respectively. They were not displayed in the experiment.

To discourage participants from adopting a predetermined strategy based on either the identity or the spatial location of the targets, both parameters varied randomly from trial to trial. Randomly sampled numbers were displayed on the vertices of either a square or a diamond, and the position of the targets was fully randomized.

Importantly, the initial and final response set sizes were not correlated. This implied that grouping trials on the basis of their initial response set size was fundamentally

different from grouping them on the basis of the final response set size. Figure 6.4 shows the number of trials per experiment for each combination of initial and final response set sizes. Blue and red groupings illustrate the differences in classifying trials on the basis of either response set. Instructed trials were selected randomly from each combination of initial and final response set sizes. Because no numbers were removed from the initial response set in instructed trials, trials with no differences between initial and final response set size (i.e., the diagonal in figure 6.4) were overrepresented in the experiment.

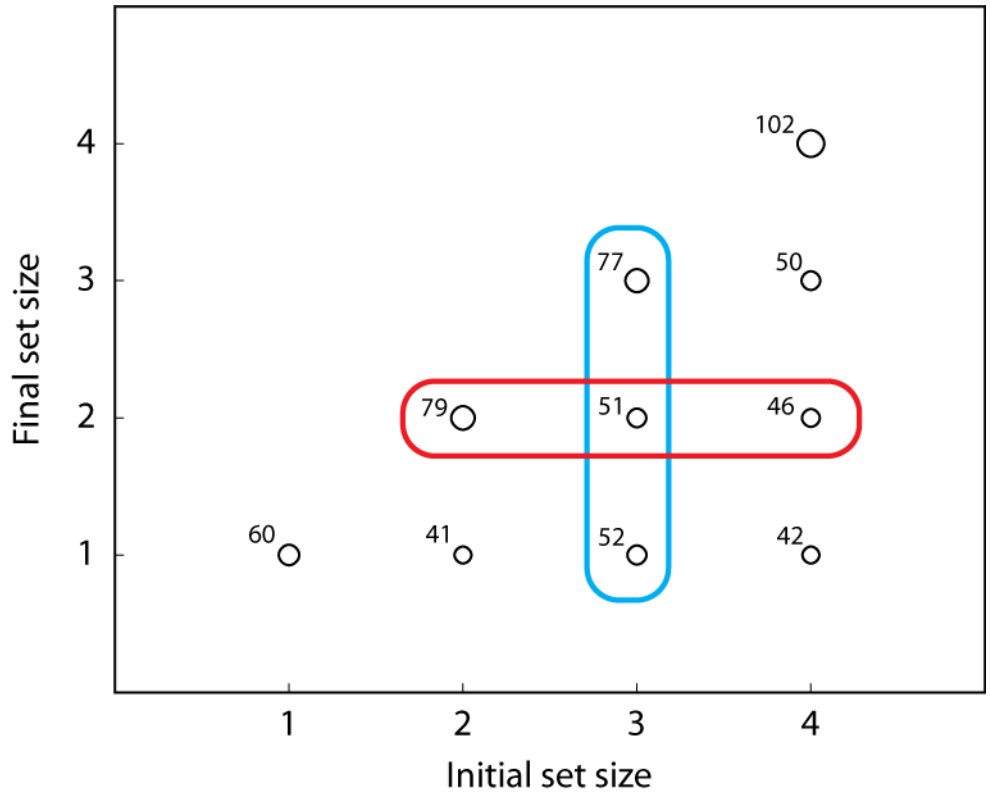


Figure 6.4 Number of trials for each combination of initial and final response set sizes. The initial response set size did not inform of the final response set size. Numbers indicate the average number of trials of each combination per participant. Blue and red groupings illustrate that the effects of each parameter (initial of final response set size) could be estimated independently. As an example, the blue group encloses the trials with an initial response set of 3. The red group encloses trials with a final response set size of 2. Because the initial and final response set sizes were not correlated, the identity of the trials was different, and led to independent datasets.

Before starting the experiment, participants had a short practice session of 40 trials. The mean movement time during this practice session was recorded to calculate the

number of rewarded trials in the first experimental block. The data from the practice session were otherwise not further analyzed.

Data analysis

RTs were calculated as the time at which the mouse position rate of change was nonzero, relative to the onset of the go signal. Because of the screen refresh rate (60Hz), RTs were obtained with a relatively low temporal precision, of one sample every 16.7 ms. Trials with RTs under 100 ms were rejected, as potentially anticipatory. In the same way, trials with RTs longer than 1000 ms were rejected. Movement times were calculated as the time taken to click within 20 pixels of the number target.

To calculate the relationship between RT and response set size, linear regressions were obtained for the data from each individual participant. The slope parameters of the linear fits for each participant were then entered into repeated measures ANOVAs where appropriate.

Hick's law (Hick, 1952) establishes that mean instructed go RTs increase as the binary logarithm of the response set size. Consequently, the typical approach to test Hick's law in a given set of responses is to include conditions with 1, 2, 4 and 8 possible response alternatives. In contrast to typical experiments testing Hick's law, this study involved the disappearance of some response alternatives from a set of 4 targets. These targets were spatially arranged. In this arrangement, response set sizes larger than 4 were problematic, because they would lead to variations in the distances between the disappeared and remaining targets, across different initial response set sizes. This reselection distance confound would be minimized in a setting with a maximum response set size of 3 (triangular arrangement). As a compromise, response set sizes of 1 to 4 possible responses were considered here. In this experiment mean RTs were fitted with a linear function, and not with a binary logarithm, for several reasons. First, there is no direct evidence that Hick's law is applicable to intentional RTs. Second, even if this were the case, the number of different response set sizes considered here (1 to 4) fall on the initial part of the binary logarithmic curve, so a linear approximation would be valid. Most

importantly, the particular shape of the functional relationship between RT and response set size was not the primary interest here. Instead, the main aim was to establish whether RTs were affected by either initial or final response set sizes, regardless the particular form of the relationship. Therefore the approach adopted was to calculate linear fits for the RTs as a function of response set sizes.

6. 3 Results

In this experiment participants were asked to select a response from an initial response set of available alternatives, and hold their response for either a short or a long SOA. At the time of the go signal, some of the initially available responses could disappear, reducing the effective response set, and yielding a smaller final response set. Participants were asked to reselect a response if their original response choice had disappeared, but to execute their original selection of it remained in the final response set.

One participant failed to understand the instructions, so their data was excluded from the analysis. This yielded a total of 17 participants. Participants made few omission errors in instructed trials. There was a mean omission rate of 0.94 ± 0.3 %. After rejection of omission trials, an average of 114 ± 2 trials were included in the instructed condition, 186 ± 5 trials in the no change condition, 126 ± 17 trials in the selected condition and 150 ± 18 trials in the reselected condition. The original selection and reselection conditions presented the highest variability in the number of trials across participants because the exact number of trials that fell in each condition depended on each participant's behaviour. Based on the total number of trials and the combination of initial and final response set sizes, the mean expected number of reselection trials was 139, which is comparable to the figure obtained.

Trials with RTs shorter than 100 ms were rejected, as potentially anticipatory. Overall, 26 ± 24 % trials were rejected, across all participants and conditions. Differences between the proportions of rejected trials were examined. A two-way 4x2 repeated-measures ANOVA with the factors of condition and SOA revealed significant differences between conditions ($F_{3,48}$=48.16, $p<0.001$). The highest

proportion of rejected trials due to anticipation was in the instructed condition. There were no significant differences between the proportion of rejected trials in the critical selected and reselected conditions ($F_{1,16}$=0.98, p=0.338). Average numbers of trials are shown in table 6.1

Table 6.1 Final mean (±SD) number of trials per condition after rejecting incorrect and anticipatory trials.

Condition	Instructed		Original selection		Reselection		No change	
SOA	Short	Long	Short	Long	Short	Long	Short	Long
Mean number of trials (± SD)	38 ± 16	38 ± 15	58 ± 22	57 ± 20	48 ± 17	47 ± 17	69 ± 22	69 ± 25

Conditions with equivalent initial and final response sets

In conditions in which no numbers disappeared from the initial response set, initial and final response sets were equivalent, so the only factors of interest were condition (instructed/no change) and SOA (short/long). The RT averaged across all participants for each response set size is shown in figure 6.5.

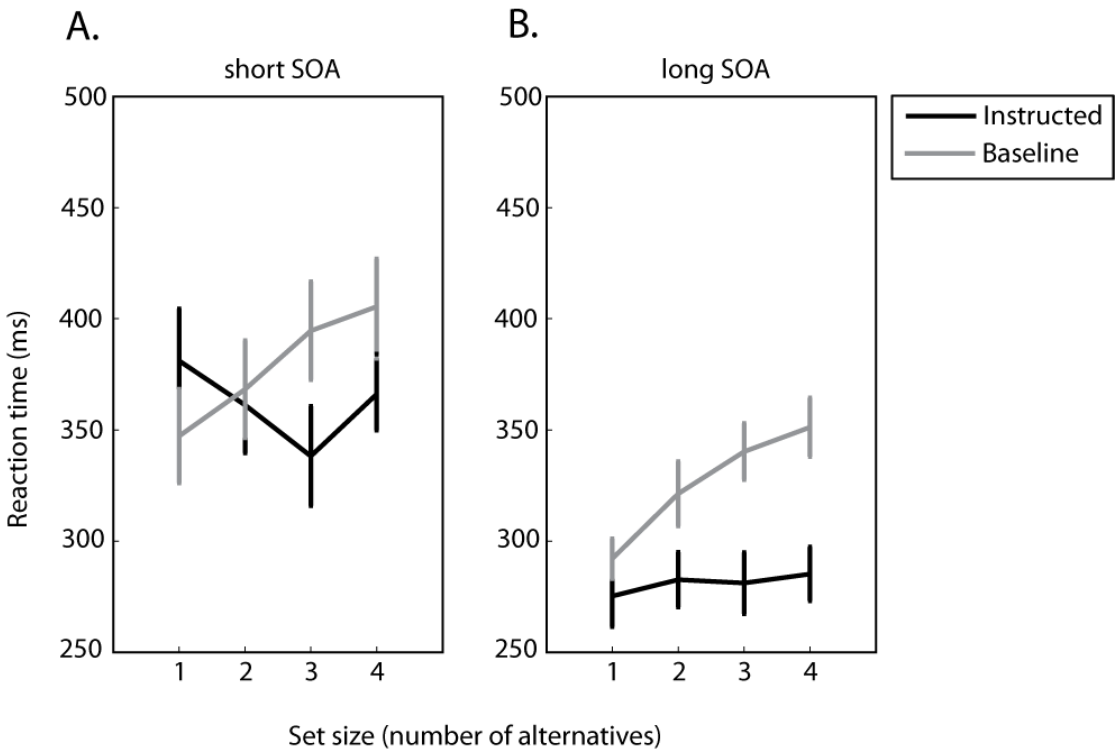


Figure 6.5 Reaction times as a function of response set size for **A.** short stimulus-onset asynchrony (SOA) trials and **B.** long SOA trials. Error bars show standard error of the mean.

To explore the effect of SOA and reselection, the mean RTs across all response set sizes were obtained for each condition (figure 6.4). A 2x2 ANOVA with the factors of condition (no change/instructed) and SOA revealed a main effect of condition ($F_{1,16}$=27.02, p<0.001), a main effect of SOA ($F_{1,16}$=22.26, p<0.001) and a significant interaction effect ($F_{1,16}$=11.01, p=0.004). Follow-up t-tests revealed significant difference for the short SOA conditions (t_{16}=-2.14, p=0.048) and a strongly significant difference for the long SOA conditions (t_{16}=-8.73, p<0.001)

These results suggest that participants prepared their motor response during the SOA, and that they achieved higher levels of action preparation in the long SOA periods than in the short SOA.

The effect of increasing response set size on RTs was further analyzed. As expected, RTs in the no change conditions trials increase monotonically with increasing response set size. This was true for both short and long SOAs (see figure 6.5). Linear

fits were obtained for the data from individual participants, and a two-way repeated measures ANOVA with the factors of condition (no change/instructed) and SOA (short/long) was performed on the estimated slope parameters.

Results revealed a main effect of condition ($F_{1,16}$=14.28, p=0.002) but no main effect of SOA ($F_{1,16}$=0.09, p=0.773) or interaction effect ($F_{1,16}$=0.14, p=0.709). The main effect of condition was expected, because intentional response selection should have an RT cost in the no change condition relative to the instructed conditions. In the latter condition, the response set size should have no effect on RT beyond the visual search time.

Conditions with different initial and final response sets

In the selected and reselected conditions, one or more numbers were removed from the response set. Consequently, the initial and final response set sizes differed. To address whether the internal representation of the response set was updated to match the new external response set, RTs were calculated as a function of each of the two possible response sets. RTs as a function of either the initial and final response set sizes are shown in figure 6.6.

Because the initial and final response set sizes were not correlated (see figure 6.4), the relationship between mean RT and the size of the initial response set was dramatically different from the relationship between the mean RT and the size of the final response set. Trials with an initial response set of 4 could have any of the possible final response set sizes of 1, 2, 3 or 4. Similarly, trials with an initial response set of 3 could have any of the possible final response set sizes of 1, 2 or 3.

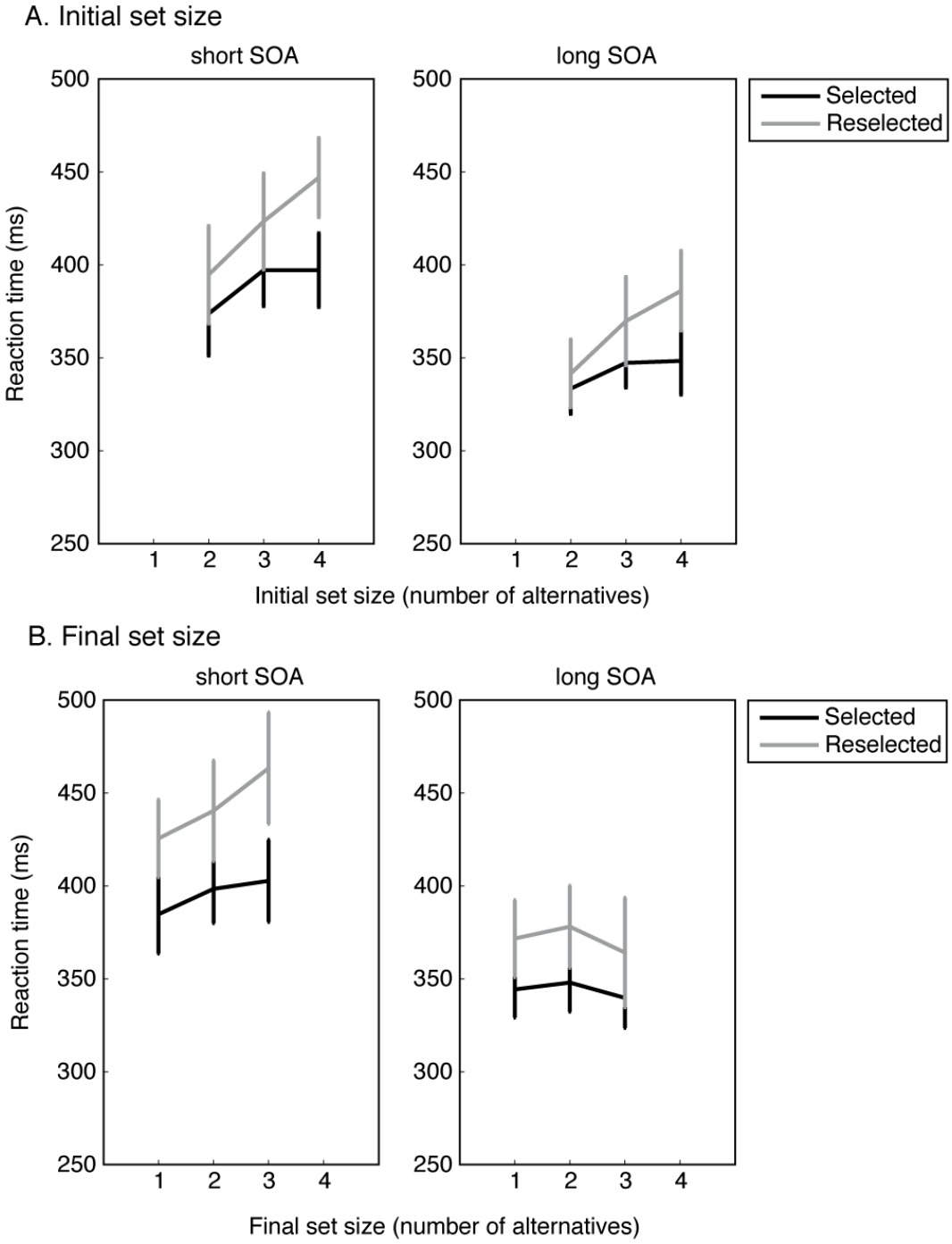


Figure 6.6 RTs averaged across all participants, as a function of the response set size for each response set (initial/final) and SOA (short/long). Panels **A** and **B** show the RTs as a function of the initial and final response set sizes, respectively. For long SOA trials, RTs increase monotonically with the initial response set size. They do not vary with increasing final response set sizes. In both conditions (selected and reselected) at least one number target disappeared from the initial response set. Therefore, there were no trials with initial response set sizes of 1 (which would have led to final response set sizes of 0) or final response set sizes of 4 (which would have required initial response set sizes larger than 4). Error bars show standard error

The initial and final response set sizes were independent (see figure 6.4). Therefore the factor of response set was incorporated into the factorial design. To examine the effects of reselection on RTs, the mean RTs were first calculated across all response set sizes.

A 2x2x2 ANOVA on the mean RTs of the selected and reselected condition revealed a main effect of response set ($F_{1,16}$=6.33, p=0.023), a main effect of SOA ($F_{1,16}$=39.24, p<0.001) and a main effect of condition ($F_{1,16}$=14.06, p=0.002). There was also a significant response set x SOA interaction ($F_{1,16}$=9.54, p=0.007). There was no response set x condition interaction ($F_{1,16}$=2.24, p=0.154), no SOA x condition interaction ($F_{1,16}$=2.27, p=0.15) and no three-way interaction ($F_{1,16}$=3.16, p=0.094).

The main effect of SOA on the mean RTs suggests that long SOAs allowed for stronger motor preparation than shorter SOAs, validating the SOA manipulation. In addition, the main effect of condition was presumably due to the RT cost of the inhibition of the original action plans and the process of number reselection.

To examine the effects of increasing response set size on RTs, the slopes for the individual linear fits of the RTs were analyzed in a repeated measures 2x2x2 ANOVA with the factors of response set (initial/final), SOA (short/long) and condition (selected/reselected). The mean slope estimates for the selected and reselected conditions are shown in figure 6.7.

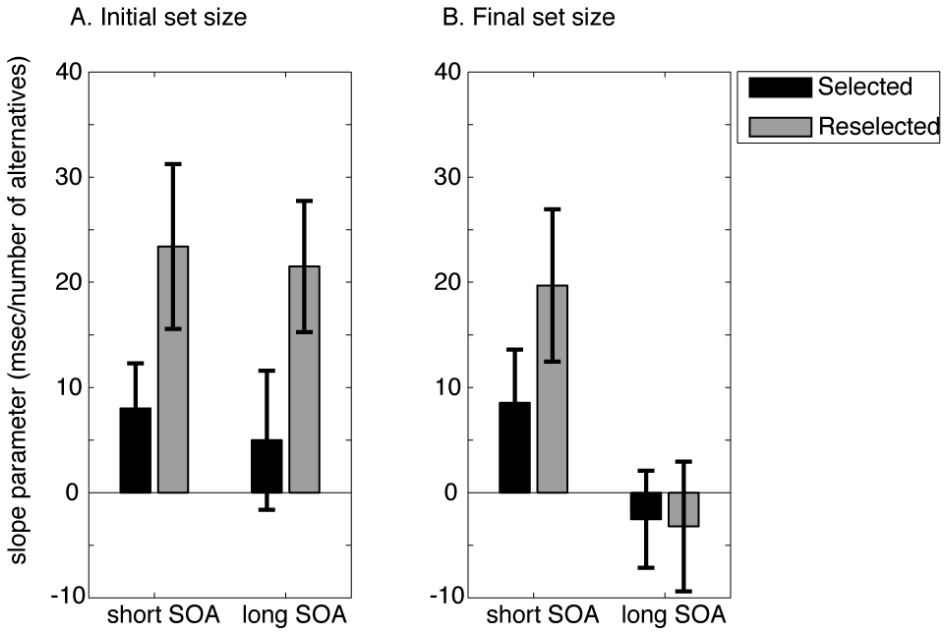


Figure 6.7 Mean slope of the linear fit to the RTs as a function of response set size, for either initial (panel **A**) or final response set sizes (panel **B**). For long SOA trials, RTs increase as a function of the initial, but not the final, response set size. Error bars show standard error.

Results from the three way ANOVA revealed a significant main effect of response set ($F_{1,16}$=5.12, p=0.038), a significant main effect of SOA ($F_{1,16}$=6.87, p=0.019) and a significant response set x SOA interaction ($F_{1,16}$=6.551, p=0.021). There was no main effect of condition ($F_{1,16}$=3.043, p=0.1), no significant response set x condition effect ($F_{1,16}$=1.61, p=0.224), no significant SOA x condition effect ($F_{1,16}$=0.305, p=0.588) and no three-way interaction effect ($F_{1,16}$=1.745, p=0.205).

The main effect of response set indicates that the initial response set size had a stronger impact on RTs than the final response set size. To investigate the response set x SOA interaction, the slope estimates were collapsed across conditions. Follow-up t-tests revealed no differences between initial and final response set sizes in the short SOA conditions (t_{16}=0.28, p=0.779), but clearly significant differences between the initial and final response set sizes in the long SOA conditions (t_{16}=4.04, p<0.001).

RTs as a function of the binary logarithm of the response set size

The above results were calculated by estimating linear fits of the RTs as a function of the different response set sizes (see Methods). As a control, it was explored whether the same results would be valid if the RTs were described as a function of the binary logarithm of response set size, as established by Hick's law (Hick, 1952). Because a maximum of 4 response set sizes are not enough to produce reliable estimates of the parameters of a logarithmic function, the response set size was linearized. The same analysis was done, but considering RTs as a function of the binary logarithm of the response set size, rather than as a function of the response set size itself. This analysis yielded similar results as the ones reported above.

A 2x2x2 ANOVA on the slopes of the RTs as a function of the binary logarithm of the response set size revealed a main effect of response set ($F_{1,16}$=9.17, p=0.008), a marginally significant effect of SOA ($F_{1,16}$=4.41, p=0.052), and a marginally significant effect of condition ($F_{1,16}$=4.41, p=0.051). There was a trend for a significant response set x condition interaction ($F_{1,16}$=3.99, p=0.06), no response set x SOA interaction ($F_{1,16}$=1.44, p=0.247) and no SOA x condition effect ($F_{1,16}$=0.16, p=0.693). There was no three-way interaction ($F_{1,16}$= 0.45, p=0.512).

Finally, RTs were calculated as the first point in time at which the speed of the cursor was nonzero. To ensure that the obtained results were not an artefact of the way in which the RTs were defined, the same analysis on the slopes of the linear fits was performed in two alternative ways. First, RTs were calculated as the time at which the cursor had covered 25% of the total distance in each trial. Second, the same analysis was performed on movement times, calculated as the time to click on the final target. In both cases, the three-way repeated-measures ANOVA yielded a significant effect of response set ($F_{1,16}$=12.8, p=0.003 and $F_{1,16}$=13.32, p=0.002, respectively)

In sum, the main effect of response set size remained after addressing the relationship between RTs and response set sizes in a way that followed more strictly the formulation of Hick's law. Importantly, this effect was not highly sensitive to the way in which the RTs were calculated.

6. 4 Discussion

This study aimed at answering the general question of whether selected response alternatives that are no longer available in the environment nevertheless remain represented in the brain. To address this question, participants were asked to intentionally select one number from an initial response set of available numbers displayed on a screen. Immediately before executing their selected response, some of the available numbers could disappear from the initial response set, yielding a final and smaller response set. This distinguished an external response set, represented by the numbers physically present in the external environment; and an internal response set, corresponding to the internal representation of the available alternatives. Using RTs as a proxy for the internal response set, this study suggests that the internal response sets driving RTs corresponded more closely to the initial than to the final external response sets. This suggests that the internal response sets are in fact resilient to external change, and "lag behind" sudden changes in the external environment.

Conditions with equivalent initial and final response set sizes: instructed and no change

In the no change and instructed conditions, no numbers were removed from the initial response set. Consequently, initial and final response sets were indistinguishable. Whereas the no change condition required intentional response selection, the instructed condition required only visual search to identify the instructed target. The no change condition is informative of the relationship between the RTs and the response set size. RTs in the no change condition showed a positive linear relation with response set size. Conversely, RTs in instructed trials did not depend on the response set size (i.e., the estimated slopes of the linear trends did not differ significantly from 0). This may seem surprising, as monotonic increases in instructed RTs as a function of response set size have been well documented (Hick, 1952). In this experiment, however, the SOA temporally separated the processes of visual search and action initiation. This may explain the null effect of response set size on instructed RTs. Importantly, this validates the SOA manipulation, aimed at

allowing for selection and motor preparation, and suggests that the results cannot be easily explained by visual search processes.

The effects of SOA (short *vs*. long) and condition (instructed *vs*. no change) were analyzed. The mean RTs across all response set sizes and the slopes of the linear fits for RTs as a function of response set size were obtained. Shorter SOAs were associated with longer mean RTs and with steeper dependencies of RTs on response set size. This suggests that longer SOAs allowed for movement preparation, reducing the mean RT and decreasing the impact of increasing the number of response alternatives.

Conditions with unequal initial and final response set sizes: original selection and reselection

In the selection and reselection conditions, some target numbers disappeared from the initial response set. Because the initial and final response set sizes were not correlated, they were incorporated as independent factors in statistical analyses. Therefore, the effects of SOA, condition (selection/reselection) and response set (initial/final) on mean RTs and on the slopes were analyzed.

Selection and reselection trials differed in two critical aspects. Reselection trials, but not simple selection trials, required inhibition of the initial response plan and reselection of a new target number. In addition, and critically, they required participants to abandon their intentionally chosen response alternative.

In both selected and reselected conditions, trials with longer SOAs showed shorter RTs. This effect mirrors what was found in the no change and instructed conditions, and once again suggests that response selection and motor preparation took place during the SOA. Further, as expected, longer RTs were found in reselection trials due to the cost of response inhibition and reselection.

Crucially, an analysis of the slopes of the linear fits revealed stronger dependencies of the RTs with initial response set sizes as compared to final response set sizes. This suggests that the initial response set size had a stronger influence on the RTs than the

final response set size. This effect was strongest particularly for long SOA conditions, as revealed by a significant response set x SOA interaction effect. Longer SOAs may allow for stronger and more stable encoding of the initial response set size, leading to more resilience of the internal representation of the initial response set.

A comparison of the selected and reselected conditions revealed that reselection processes had a statistically significant effect on the mean RTs, collapsed across all response set sizes. Interestingly however, reselection did not have a clear effect on the RT slopes. RTs increased with the initial response set size in both selected and reselected trials. There was a numerical difference between the estimated slopes for selection and reselection trials, but the differences were not statistically significant. The factor of condition was not associated with any significant interaction effects with other factors. Therefore, the persisting influence of the initial response set size does not appear to be related to the intentional selection of response alternatives. Instead, results suggest that it is the nonspecific encoding of the response set that makes it resilient to external change.

Intriguingly, a marginally significant effect of condition on the RT slopes was found when the binary logarithm of the response set size was considered instead of the absolute response set size. This analysis was motivated by exploring a strict implementation of Hick's law, which establishes that instructed go RTs vary linearly with the binary logarithm of the response set. However, there is no solid empirical evidence for such a strict implementation of Hick's law, so the potential effects of intentional selection remain speculative.

Additional controls showed that the significant effect of response set size is not an artefact of the way in which RTs were measured. Two additional controls considered complete movement times, or measured RTs as the time at which the distance travelled by the cursor was 25% of the final distance. In both cases, a statistically significant effect of response set size was found.

In light of the results from chapters 2 and 3 in this thesis, it might be hypothesised that instructed decisions would lead to even stronger and more resilient internal

representations of the response space. The comparison between intentional and instructed action selection could not be done here, because no numbers were removed from the initial response set in instructed trials. Speculatively, if the instructed target was removed from the initial response set, the numerical differences in the slopes between instructed-selected and instructed-reselected trials may become statistically significant.

Neural representations of response sets

The behavioural analyses reported here do not inform about the nature of the internal representations of response sets. Data from monkey electrophysiology are especially relevant to this issue.

Cisek and Kalaska (Cisek & Kalaska, 2005) showed monkeys a screen displaying two potential targets of a saccade reaching task. Each target fell in the receptive field of different neurons. After a given delay, monkeys saw a colour cue specifying which of the two initial potential targets they should saccade to. The authors recorded electrophysiological activity from single cells in PMd. They found a large proportion of the neurons that presented sustained firing rates during the delay period if either of the two potential targets fell in their receptive fields. Once the colour cue specified which was the correct target, these "potential-response cells" rapidly increased their firing rate for the correct target, and decreased the firing rate for the incorrect target. This strongly suggests that potential responses are represented internally during an action anticipation period.

Further, Cisek and Kalaska compared the overall firing rates for delay periods with one and two potential targets. They found that the overall population neural activity for two potential targets was lower than the sum of the neural activity for the two targets presented separately. This suggests possible competitive interactions between the representations of available responses.

Accumulator models for response selection

In agreement with these results, Purcell *et al* (Purcell, Schall, Logan, & Palmeri, 2012) found neural activity correlating with response set sizes. The authors trained monkeys to saccade to an instructed target amongst a group of distractors. Neuronal populations in both visual areas (frontal eye field -FEF-, superior colliculus -SC- and lateral intraparietal area -LIP-) and motor areas showed increasingly lower firing rates with increasingly larger response sets (2, 4 or 8 total items). Firing rate in neuronal populations is normally taken to be necessary for perceptual accumulator models (Gold & Shadlen, 2001; P. L. Smith & Ratcliff, 2004). Broadly these models hold that once enough "perceptual evidence" (i.e., neural firing) is accumulated in a neuronal ensemble representing a response alternative the corresponding sensory and integrative visual areas will project to motor areas to drive the corresponding action.

The results from Purcell *et al* (2012) provide a plausible neural explanation for Hick's law. Larger response set sizes will response set a lower baseline firing rate from which perceptual evidence needs to be accumulated until it reaches a decision threshold. Lower baseline levels will require more accumulation of information to reach the decision threshold. In turn, this may translate into longer accumulation times, manifest as longer RTs.

These accumulator models had traditionally been restricted to perceptual decision-making, where the accumulation of neural evidence comes from the sensory information about the external environment. Recently however, Zhang et al (Zhang et al., 2012) have adapted the accumulator models to intentional choices in human behaviour. The authors asked participants to make either intentional or instructed finger tapping movements. Based on the behavioural data, they estimated the parameters for the optimal accumulator model on a trial-by-trial basis. Interestingly, when the authors correlated these estimates with blood-oxygen level dependent (BOLD) signal data from the whole brain, they found significant correlations between expected relative neural activity and BOLD signal levels in areas that overlap with those that have been typically associated with intentional action (namely supplementary motor area, -SMA-, preSMA, anterior cingulate cortex -ACC- and sensory cortex).

These, "intentional" accumulator models, analogous to perceptual models, may relate to the present findings. Different neuronal assemblies may gather "intentional" information for each target. These assemblies may provide the neural implementation of the internal response sets of this study. Speculatively, spiking activity in neuronal assemblies that correspond to the alternatives that are no longer available may not be fully inhibited immediately after target disappearance. In line with the results reported by Purcell *et al* (2012), initial firing rates of each neural assembly may be lower for larger initial response set sizes in this experiment, leading to longer intentional RTs.

PFC function in the "sculpting" of the response set

In two experiments, Fletcher *et al* (Fletcher, Shallice, & Dolan, 2000) used fMRI to examine the role of the PFC during memory encoding. In the first experiment, participants were asked to learn word pair associations in two stages. In the first stage, they were asked to memorize novel word pairs. The second stage provided the experimental manipulation. In the "new pairs" condition, participants were simply asked to memorize a second novel response set of word pairs. In the crucial "re-paired" condition, participants were asked to remember a new list of word pairs that contained items from the previously learned word pair list. In this condition, participants had to inhibit the word pair associations learnt in the first stage, in order to re-pair them according to the new response set. In this way, BOLD signal activity associated with the contrast between the re-paired and the new-pairs condition would be related to the inhibition of the word pair associations that had been learnt in the first stage.

In the second experiment, participants were asked to memorize either closely related word pairs (e.g., "King, Queen") or distantly related word pairs ("Prince, Skull"). In this case, increased BOLD signal activity for the contrast between distantly and closely related word pairs would indicate the positive word association process.

Results revealed increased BOLD signal levels in the left ventrolateral PFC (VLPFC) for both contrasts, in cases of facilitation and inhibition of links between concepts. On the basis of these results, the authors suggest that the VLPFC plays a

role in "sculpting" the response set. In other words, VLPFC may be playing the dual role of associating the necessary concepts and semantic terms necessary for the word pairing association, and at the same time inhibiting those responses that are no longer relevant for the task and would otherwise interfere with task performance.

Reward-based decision making paradigms have shown that the relative value of available responses is represented in the PFC. In one study, Boorman *et al* (Boorman, Behrens, Woolrich, & Rushworth, 2009) measured brain BOLD activity while participants performed a simple reward-based decision-making task, where they were asked to choose between two response alternatives. The probability of reward of each response alternative changed over time and depended on the recent trials history. In this way, participants would typically repeatedly choose one response alternative until there was enough evidence to favour the alternative response. Once this occurred, participants would switch to a new series of repeated choices of the second alternative. Interestingly, Boorman *et al* found that BOLD activity in the frontopolar cortex (FPC) correlated with the expected value of the counterfactual alternative. Further, immediately before a switch event, there was increased functional connectivity between FPC and parietal and premotor regions, suggesting that FPC triggered the switch to a new response once enough evidence in favour of the alternative response had accumulated.

Together, these results provide a plausible implementation of the internal representations of the response alternatives. Activity in neuronal ensembles in premotor and parietal sensorimotor areas may code for the response alternatives. Prefrontal function may drive activity in these areas by top-down regulatory mechanisms. Intentional decisions may maintain these representations relatively more active than instructed decisions.

BOLD correlates of response sets

In human participants, neuroimaging experiments have aimed to identify areas where BOLD signal activity correlates with response set sizes. For example van Eimeren *et al* (van Eimeren et al., 2006) have measured BOLD signal activity while participants intentionally chose one out of a response set of 1 to 4 possible responses. They found

that ACC, SMA and PMd showed increased BOLD activity when intentional selection was required, but the increases did not depend on the number of possible alternatives.

Role of internal representation of response alternatives in response selection

Traditional models of response selection had assumed that the processes of perception, response selection, and response execution are temporally segregated e.g., (Bhushan & Shadmehr, 1999; Flash & Hogan, 1985). However, more recent theoretical (Cisek, 2007) and empirical (Klein-Flügge & Bestmann, 2012) accounts suggest that response selection and action preparation occur in parallel. By this account, the internal representation of the available alternatives is not an isolated process, but is instead intimately linked to response selection. Cisek (2007) has put forward the "affordance competition hypothesis". This hypothesis suggests that overt behaviour is the result of the competition between potential responses, and the potential action plans required to make them. Klein-Flügge and Bestmann (2012) have recently provided evidence supporting the affordance competition hypothesis. In a recent TMS study, they asked participants to choose between two targets with either their right or left hands. The right and left targets would be associated with different reward probabilities, biasing the participants' choices. The authors showed that costicospinal excitability (a measure of motor preparation) showed lateralization even before the response selection process had completed, providing evidence consistent with parallel rather than serial processing for response selection and execution.

Limitations of this study

Several limitations should be considered here. First, and most importantly, the results of this study rely critically on the validity of the assumed link between RT and internal set size. It was assumed that go RTs increased monotonically with the number of items in the internal representation of the response set size. This assumption was based on a generalized version of Hick's law, and was as such

supported by empirical studies on instructed action (Hick, 1952) and intentional action (H. Lau, Rogers, Ramnani, et al., 2004; van Eimeren et al., 2006). In addition, this assumption was supported by the data from the RTs in baseline and instructed trials. However, on the basis of behavioural data alone, it cannot be further validated.

Second, these results present an important limitation in the context of this thesis. They describe the "persistence of the internal representation of non-selected alternatives in cases of intentional selection". However, the experimental design did not allow for a direct comparison between intentional and instructed conditions. Therefore, it cannot be assumed that this persistence effect is exclusive to intentional conditions; nor do these results reveal the specific features of intentional control as opposed to other forms of action control.

A third limitation of the study is related to the reliance on subjective report. Trials were identified as selected or reselected on the basis of subjective report, which cannot be checked objectively by an external observer. Subjective report should always be treated with caution, precisely because there is no objective way to confirm its relation to the actual mental processes taking place. However, in this task, if participants' subjective report was not truthful, there would have been no differences in mean RTs between the selected and reselected conditions. Instead, longer RTs were found for reselected than for selected conditions, arguing for the validity of the subjectively-based sorting strategy. In addition, inaccurate subjective report would have only reduced any possible significant differences.

Finally, this study did not address action inhibition directly, but instead assumed it had occurred during trials where participants reported they had reselected a response. Under this paradigm, therefore, the processes of action inhibition and response reselection are difficult to segregate and separately characterize.

Interestingly, in a review paper, Mostofsky and Simmonds (Mostofsky & Simmonds, 2008) have suggested that response inhibition and response selection are not distinct processes, but are instead "two sides of the same coin". Mostofsky and Simmonds reviewed results from monkey electrophysiology and behavioural and neuroimaging studies on human participants. The authors note that in a typical go/nogo task, a nogo response requires an active decision not to move. In other words, a successful

nogo trial requires both inhibition of the prepotent go response, and response reselection processes. In that way, processes of response inhibition and response reselection are not distinguishable in a typical go/nogo task. Importantly, they note that the precise neural processes and correlates of response selection and inhibition may depend on the specific task demands. The present task may be one in which response inhibition and reselection cannot be distinguished.

Implications of this study

A 17th century philosophical debate, on the definition of freedom of choice, resonates with the situation studied here (Hobbes & Bramhall, 1999). In this debate, between Thomas Hobbes and John Bramhall, Hobbes held that free choice is that made in "…the absence of external impediments" (Hobbes, 1937). Bramhall disagreed, arguing that Hobbes' definition was ambiguous. He imagined a man that makes a decision, unaware of external impediments. Bramhall believed that, according to Hobbes' definition, the outcome of the man's intentional decision would determine whether he was free or not, at the time of decision. He had been free if he chose the course of action that was possible given the external circumstances; and he had not been free otherwise (Bramhall, 1655). This poses a problem because freedom of a decision should be linked to the decision itself, at the time at which is occurs. Bramhall argues agains Hobbes' definition from the absurd, showing that if the relative freedom of a decision can be retrospectively changed by the decision outcome, a person could simultaneously be "free and not free".

Four hundred years had now passed since the debate between Hobbes and Bramhall. The results presented here offer an answer to their questions. They suggest that if a man has decided not to play tennis, and then finds the tennis court closed, his freedom will still be hindered. This is because his internal representation of the response alternatives may still contain the possibility of playing tennis, even after having intentionally decided not to do so and even after being aware that he cannot.

Future directions

Two interesting questions arise from these results that may be explored in future studies. First, the important issue of differences between instructed and intentional trials could be addressed. Here, the task was designed to emphasize the validity of instructed trials. Removing the instructed target from the initial response set would have reduced the strength of the instructions. Consequently, no numbers were removed from the initial response sets in instructed trials. In a follow-up experiment, a minority of instructed trials with disappearing targets could be incorporated into the design, to compare internal response set representations across instructed and intentional conditions.

Second, these results suggest that the internal response set size is not immediately "re-sculpted" to match changes in the external environment. Presumably however, the internal and external response sets will match given sufficient time. The time course of the re-sculpting of the internal response set could be addressed in future studies.

6. 5 Conclusion

Here, the internal representation of response alternatives was investigated. The results revealed that once an internal representation has been established, it is relatively resistant to change. That is, response alternatives that suddenly become unavailable in the external environment may nevertheless still be actively represented, and may still affect behaviour. This experiment did not allow for a direct comparison between the flexibility of the internal representation of response spaces in cases of intentional selection and cases of instructed selection. However, these results suggest an interesting flip side of the relative weakness of intentional decisions. Weaker intentional decisions are consistent with the persistence of the internal representation of the counterfactual alternatives. Neurophysiological data

suggest that in cases of active maintenance of multiple response alternatives (in this case, selected and non-selected), all representations are scaled down. Therefore, in cases in which the alternative representations are not suppressed, the selected representation is less salient.

Chapter 7 Probing new methods to measure the time of conscious intentions

The time of conscious intentions to act has previously been measured by retrospective methods, famously implemented by Libet and colleagues. The experiment described in this chapter implemented a recently proposed improvement to the heavily criticized retrospective methods to measure time of awareness of intentions. In particular, it was addressed whether the online method could be used to reveal potential differences between the times of "onset of awareness" between instructed and intentional actions. These putative differences could in turn be related to differential levels of action preparation in the two types of action. A systematic analysis of the method revealed some of the weaknesses that may render it unsuitable for detecting subtle differences between conditions

7. 1 Introduction

Methods of mental chronometry provide an important way to measure and then make inferences about mental processes underlying behaviour (Posner, 2005). In particular, mental chronometry offers a substantial set of tools for investigating the relation between the common experience of action, relative to both the physical source event itself, and the neural events that encode it.

Mental chronometry methods to measure time of awareness of intention

Libet et al (Libet et al., 1983) famously used mental chronometry to make psychophysiological inferences about the causes of voluntary actions. Libet et al found that awareness of an impending action followed, rather than preceded, measurable neural activity that signals motor preparation. Therefore they argued that conscious thought could not possibly have a causal role in movement initiation, or decisions about when to make actions.

Time of awareness of action selection

Using the same principle of mental chronometry, Haggard and Eimer (Haggard & Eimer, 1999) addressed the relationship between neural activity and decisions about what action to make. The authors examined whether the time of conscious intentions was related to the onset time of the associated RPs. They measured EEG activity in 8 participants in a Libet task, making self-paced key presses and retrospectively reporting the time at which they had become aware of their intentions to make key presses. The authors extended the Libet task by including two different conditions. In the instructed condition, participants were instructed on which hand to use throughout the block. In the intentional condition, participants were asked to freely choose the effector hand in every trial. Haggard and Eimer failed to replicate previous findings by Praamstra et al (Praamstra et al., 1995) and Dirnberger et al (Dirnberger et al., 1998), as they found no differences in the amplitude or onset of the RP or LRP between the intentional and instructed conditions. Haggard and Eimer then compared the LRP onset times of trials with early *vs*. late judgements of

conscious intentions -W judgements (Libet et al., 1983)-. They found that trials with early W judgements were associated with earlier LRP onset times than trials with late W judgements On the basis of the observed direct relationship between LRP onset and W judgements, the authors suggest that LRP may have a causal role in awareness of motor intention.

In the case of Haggard and Eimer's study, the assessment of the W-time was estimated using the classical retrospective judgement first implemented by Libet *et al*. This method addresses the time of conscious intentions directly, by asking participants to recall and report the position of a clock hand at the time they first became aware of their motor intentions. Retrospective methods remain a common means to determine onset of subjective awareness of intention (e.g., Soon, Brass, Heinze, & Haynes, 2008), though their use is controversial, as it will be discussed below.

Methodological and conceptual problems with the retrospective method

The retrospective method has been widely discussed (Gomes, 2002; Joordens, van Duijn, & Spalek, 2002; Libet, 1985, 2000). It has been criticized mainly on the basis of potential temporal biases in making W judgements. Because the method relies on retrospective reconstruction, and because introspection about W-time is unusual and difficult, it is in principle possible that subjects do not report *bona fide* percepts, but simply what they think they are expected to feel. If this were the case, W judgements would reflect folk knowledge rather than true timing of "conscious intentions". Even If there is a genuine percept of conscious intention, several factors, notably the allocation of attention, may bias the judgement of their time of occurrence. The retrospective method has been used to argue strongly against a causal role for conscious intentions on motor behaviour. The conclusions from these studies rely on relatively large effect sizes (250 ms in the case of Libet *et al*, 1983 and 8 s in the case of Soon et al, 2008). As such, they are not challenged by the biases of under approximately 100 ms that retrospective methods might introduce (Joordens et al., 2002). However, in the experiment conducted by Haggard and Eimer, potential differences in the time of conscious intentions between instructed and intentional actions may be very small, compared to the relatively large retrospective biases. This

may be one reason why Haggard and Eimer found no differences in the reported W-time between instructed and intentional movements.

Online method to measure time of conscious motor intentions

Matsuhashi and Hallett (Matsuhashi & Hallett, 2008) have recently suggested an elegant method to asses conscious intentions that works around the problem of potential retrospective biases. The method probes conscious intentions *online* rather than retrospectively. Briefly, in the online method participants are asked to make self-paced movements, and to be mindful of their intentions to move. While participants engage in this task, tones are played at random times. Participants are then asked to ignore any tones that occur at times when they were not about to move, and continue with the task. However, if a tone occurred when participants were aware of preparing to move, they should inhibit their movement. In this way, participants' otherwise uniform rate of movement is only interrupted by tones occurring during periods of intention awareness. Consequently, the timing of the externally controlled tones serves a marker of two possible mental states (namely, aware *vs*. not aware of intentions), which can be classified on the basis of the behavioural output following each tone.

Importantly, Matsuhashi and Hallet found that the online method yielded estimates of conscious intentions that were about 1 s earlier than the typical values of W-judgement found in studies using the Libet task. Matsuhashi and Hallet argue that this difference may be related to the explicit probing method in the online task. Recalling Haggard and Eimer's result, they found that the LRP preceded the W-judgement by 0.5 s, which was consistent with a causal role of LRPs in conscious intentions. But if an online method that does not suffer from retrospective biases brings the timing of conscious intention earlier than the LRP onset, this causal role may be questioned. Therefore the issue of timing of conscious intentions remains of great interest and would benefit from bias-free approaches.

Online method to measure time of awareness in case of intentional response selection

Because the online method does not depend on potentially large retrospective biases, it may be suited to explore fine-grained differences in the timing of awareness between two closely related conditions. The aim of this study was therefore twofold. First, it aimed at exploring the validity and potential uses of the online method to determine its sensitivity. Second, it aimed at applying the online method to explore whether there are any differences between the timing of conscious intentions in conditions of instructed and intentional action. The hypothesis was that participants would show earlier times of intention for intentional actions as compared to instructed actions.

Here, participants were instructed to inhibit action when an external tone fell during the period of intention awareness. Importantly however, this study was not aimed at studying inhibition of action directly. Inhibition was used only as a tool to indirectly measure times of onset of conscious experience.

7. 2 Methods

Participants

Sixteen naïve healthy volunteers (6 female, 2 left handed; mean age ± SD 24 ± 6 years) participated in this study.

Task and stimuli

The experimental design closely followed that described by Matsuhashi and Hallett (Matsuhashi & Hallett, 2008). Participants sat with their index fingers each resting on a force-sensitive resistor (Active Robots Ltd, Somerset, UK). Analogue signals from the sensors were recorded through a data acquisition device (USB 6008, National Instruments, Berkshire, UK).

Participants were asked to make sudden right or left index finger lift movements when they felt the urge to do so. They were asked to make movements at an

approximate rate of 1 movement every 5 -10 s, and to direct their attention towards their motor intentions. It was stressed that they should make a movement as soon as they became aware of their intentions to move, and to avoid deliberately introducing delays between the time of awareness and the time of action.

While participants were making movements, 1000 Hz tones were played through loudspeakers. The timing of the tones was determined at the beginning of the experiment, and was therefore independent of the participants' behaviour. The interval between tones was sampled from a uniform distribution with a minimum of 2 s and a maximum of 5 s. Participants were asked to ignore any tones that occurred at times when they were not aware of their intentions to move. However, if a tone occurred when participants were aware of being preparing a movement, they were asked to inhibit their action plans, and wait for a minimum "reset period" of 5 s before planning the next movement. In an *a posteriori* analysis, each tone was associated with a measure of the temporal difference between that tone and the nearest movement within a given time window (see below). In this way, tones that occurred *before* the emergence of conscious motor intentions would be followed by a movement shortly after. However, if tones occurred after the onset of conscious motor intentions, the prepared movements would be inhibited and the temporal difference between the tone leading to inhibition and the subsequent movement would be longer, as they would be separated by (at least) the duration of the "reset period" of 5s (see figure 7.1A).

In this way, the task provides a means to estimate the time of conscious intention onset in an *online* fashion, without resort to the retrospective judgements that are so problematic in the Libet task (see above). Tones that fall within the period of conscious awareness will lead to movements being inhibited (see figure 7.1 B). In consequence, a profile of the distribution of tones occurring before the time of movement should show a decrease in the tone frequency just before the movement (see figure 7.1C), because those tones leading to movement inhibition will have been effectively "removed" from the distribution (see figure 7.1 E). Tones occurring *immediately* before the movement will also contribute to the distribution as they will fall after the "point of no return" (i.e., so close to the impending movement that it cannot be stopped cf. the stop signal reaction time, see figure 7.1D), (de Jong, Coles,

Logan, & Gratton, 1990), and action inhibition processes may not be quick enough to achieve inhibition.

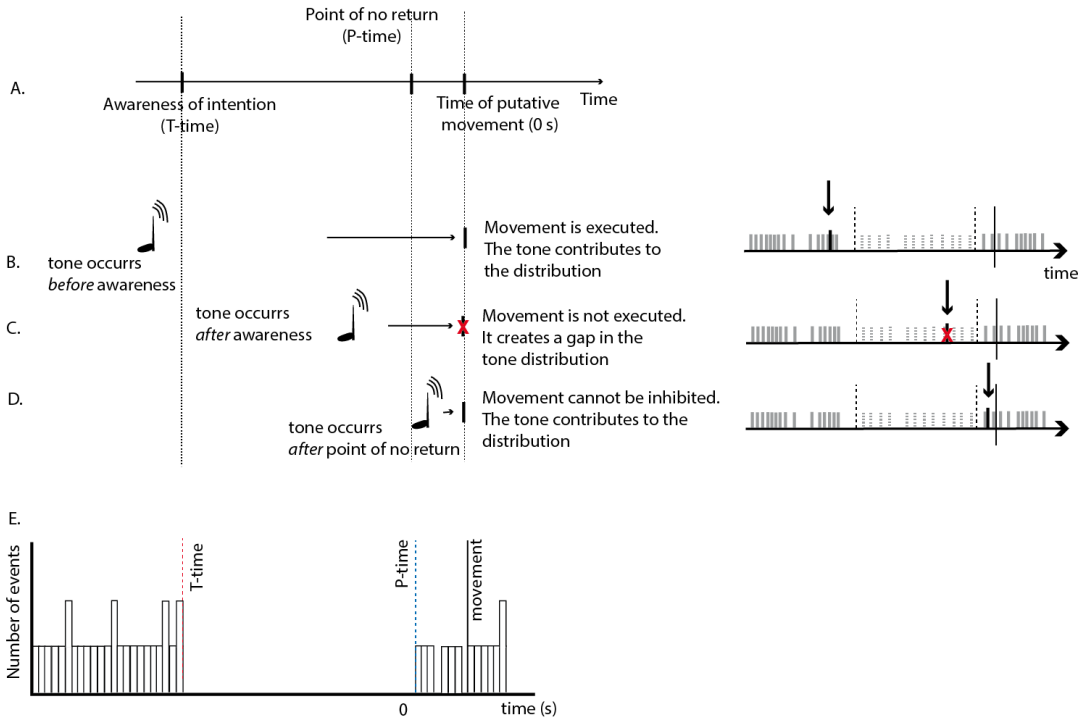


Figure 7.1 Task schematic. **A.** Timing of events. Participants become aware of their intentions (T-time) and given amount of time before they make their actions. The time of tone relative to conscious awareness determines whether movements occur or not (**B** through **D**). **B**. Tones occurring *before* conscious awareness of intention are ignored. If the time difference between the tone and the subsequent movement is shorter than 5 s (see text), the time difference is computed and considered for the distribution. **C**. Tones occurring *after* the time of awareness of intentions signal that movements should be inhibited. The time difference between tones leading to inhibition and the inhibited movement is not detected, and generates a gap in the tone distribution. **D**. Tones that occur after the time of awareness but too close to the movement to lead to timely action inhibition (i.e., tones that fall after the point of no return) contribute to the tone distribution. **E** A putative example resulting dataset. Histograms show the tone frequency distribution, relative to movement onset. T-time indicates the time of conscious intention, and is associated with a decrease in the tone frequency distribution. P-time indicates the point of no return, and is associated with an increase in the tone frequency distribution.

In line with Matsuhashi and Hallet's (2008) terminology, *T-time* refers to the time of awareness of intention; and *P-time* refers to the point of no return.

To test whether the time of conscious awareness differs between the instructed and intentional conditions, participants were asked to make movements in one of two conditions. In the instructed blocks (6 blocks in total, 3 with each hand), the required effector finger was indicated at the beginning of each block and kept fixed throughout. In the intentional conditions (6 blocks in total), participants were told to decide before each movement which finger they would use. They were asked not to pre-decide which finger they would use, but instead to decide at the last moment, when they were about to make the movement.

The experiment was divided in 12 blocks, and lasted for approximately 50 min. The order of the blocks was randomized across participants. There were no individualized trials, but each block was comprised of 35 tones. The exact number of movements per block depended on each participants' behaviour.

In addition, in order to monitor behaviour, 14 participants were asked to report verbally whenever they had inhibited an action. Their verbal reports were recorded with the audio editing software Audacity (http://audacity.sourceforge.net).

Data analysis

1. Parametric fit of error function

Each tone was associated *a posteriori* with its time of occurrence, relative to the nearest movement. Each movement was therefore associated with all tones that fell within a time window of -5 s to 2 s. A tone distribution was then calculated, by combining all tones time-locked to the nearest movement within the selected time window.

Following Matsuhashi and Hallett, the time of intention was estimated in two ways; namely parametrically, fitting a sigmoid curve, and nonparametrically, by calculating a density function on the basis of the data.

In the parametric approach a cumulative normal function (*erf*) was fitted to the data (equation 7.1).

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2}\, dt \tag{7.1}$$

To estimate T-time, all tones occurring at any time between -5 s and -0.01 s before movement were considered, and the remaining time (-0.01 to 2 s after movement) was padded with zeros. An *erf* function that fits to the expected profile for the distributions of tones for the T period (i.e., descending) will follow equation 7.2.

$$F(x) = p_1 \left\{ 1 - erf\left(\frac{x - p2}{\sqrt{2p_3}} \right) \right\} \tag{7.2}$$

Three parameters (p_1, p_2 and p_3) are enough to describe the shape of the error function. The effect of varying the value of each of these parameters is illustrated in figure 7.2. p_1 shifts the curve vertically; p_2 shifts the curve horizontally and p_3 determines the slope of the transition. If the time of conscious intention differs between conditions, then the position along the x axis of the step in the error function should differ. The final estimates for p_2 were therefore subjected to statistical analyses.

In turn, parameters p_1, p_2 and p_3 reflect different cognitive processes. p_1 increases with an increased overall tendency to move. Impulsive, hyperactive participants might therefore be expected to produce higher p_1 values. p_2 represents the onset time of conscious intentions to act. In this experiment, p_2 is expected to be negative, as conscious intentions are expected to precede action in the healthy population. More negative values of p_2 (values further away from 0) would indicate earlier awareness of intention, relative to the time of action. The final parameter p_3 indicates the consistency of the timing at which conscious motor intentions emerge, relative to

movement onset. p_3 may also be taken to represent precision with which the intention is perceived. If the conscious intention to act emerges at a reliable time relative to movement, this will give rise to a very steep fall in the tone histogram. Steep falls are associated low p_3 values.

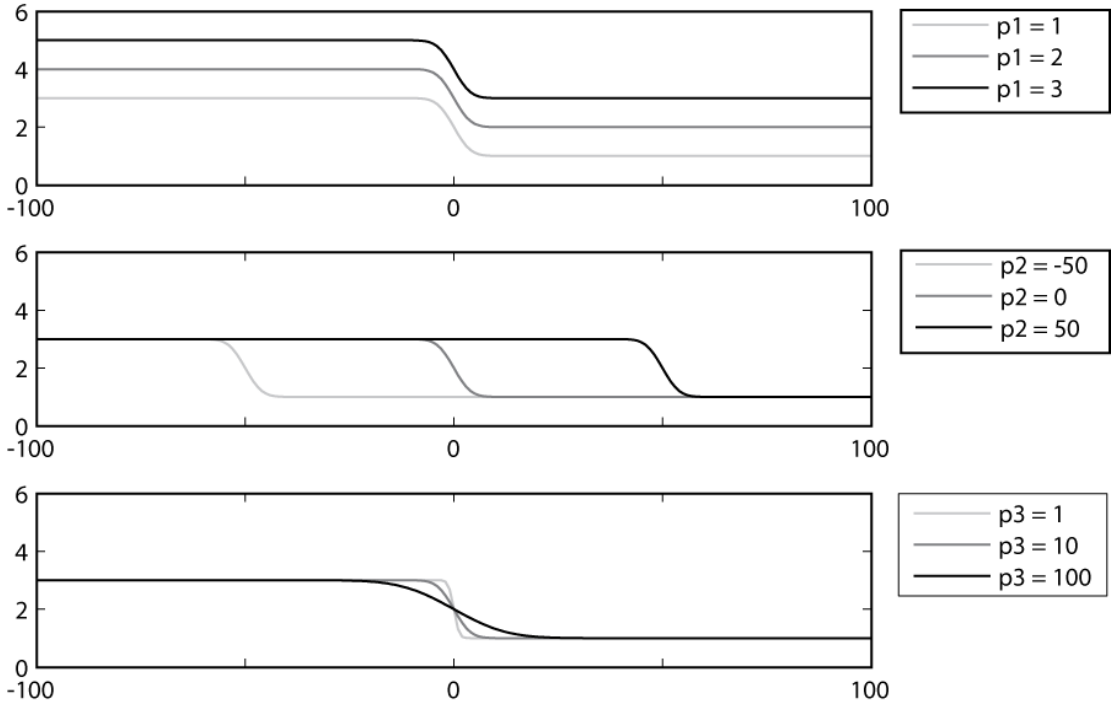


Figure 7.**2** Effect of varying parameters p_1, p_2 and p_3 on the shape of the error function used to describe the experimental data. **A.** p_1 shifts the curve vertically; higher p_1 values indicate higher overall movement frequency **B.** p_2 shifts the curve horizontally. More negative p_2 values indicate earlier intention onsets **C.** p_3 gives the slope of the transition. Smaller p_1 values indicate more precision and temporal reliability of the onset of conscious intentions relative to the time of movement.

2. Nonparametric estimation of density function

A second possible method to estimate T-time is by following a nonparametric approach and estimating a density function from the tone distribution. A classical method for estimating the density function is the kernel density estimation (Parzen, 1962; Rosenblatt, 1956). Here, a discrete distribution may be convolved with a kernel function (such as a gaussian kernel), to obtain a smoothed distribution in continuous space. The advantage of this method over the parametric fit to the *erf* function is that whereas the latter depends on the selection of the bin size and bin centres for the initial event frequency estimation, the density estimation does not.

However, density estimation methods are left with one free parameter, namely the bandwidth of the gaussian kernel, which will influence the goodness of fit of the density function to the underlying data. Smaller bandwidths allow for more precise fits to the data, at the expense of increased complexity of the resulting density function (see figure 7.3).

The density function *f(x)* is calculated at each point in time with equation 7.3,

$$f(x) = \sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - X_i}{h_i}\right) \tag{7.3}$$

where x is the time relative to the movement onset, and X_i is the timing of the i-th tone. The density function f at each x is given by a weighted sum of all tones. The Gaussian kernel K (equation 7.4) determines the contribution of each Xi to the amplitude of the density function in x; as a function of the temporal distance between X and X_i (equation 7.3); and of a variable bandwidth h_i (equation 7.5). From equations 7.4 and 7.5 it follows that the greater the temporal difference X-Xi, the smaller the amplitude of the kernel. In this way, events that lie temporally far away from x will have little impact on the amplitude of *f(x)*.

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \tag{7.4}$$

In addition, each Gaussian kernel is modulated not only on its amplitude but also on its bandwidth. The variable bandwidth is obtained by modulating a global initial bandwidth h_0 with a factor that recursively depends on the density function at each X_i.

$$h_i = h_0 f_p^{-\frac{1}{2}}(X_i) \tag{7.5}$$

Following the method adopted by Matsuhashi and Hallet, the final density function *f(x)* was obtained in two iterative steps. An initial pilot function f_p was calculated as the overall average of tones across the whole analysis period (-5 s to -0.1 s). The density function obtained was used as a pilot function in a second iterative step. This process was then repeated to obtain the final density function.

From equation 5 it follows that periods with denser tone frequency will be associated with a narrower bandwidth for the following iteration in the estimation of the density function. Instead, periods with sparser tone distributions will have wider bandwidths, allowing for a good compromise between an efficient smoothing and a precise fit. Figure 7.3 illustrates the dependency of the estimated density function on the choice of initial bandwidth.

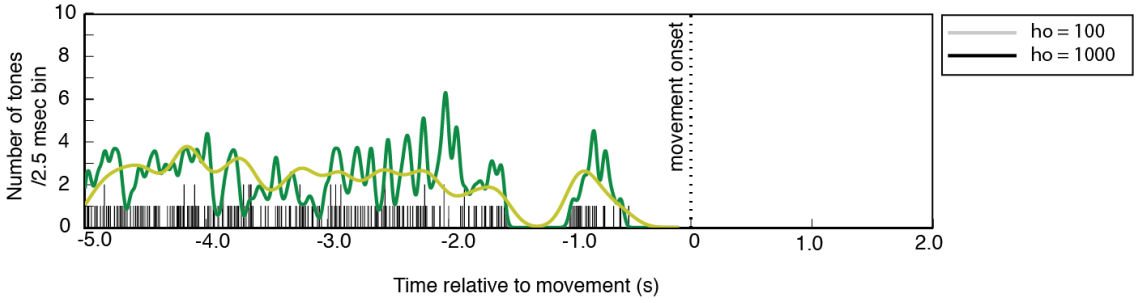


Figure 7.3 Effect of varying the parameters used to find the density function to describe a tone distribution. Simulated data illustrates the effect of varying the initial bandwidth parameter h_o. Smaller h_o decrease the overall smoothness of the density function, and consequently increase the goodness of fit of the density function.

T-time was estimated via bootstrapping. 1000 independent subsamples were drawn with replacement from the original time distribution. Subsamples were of the same size as the original distribution. The entire iterative estimation of the density function was then repeated 1000 times, for each subsample. The mean value of the density function was then obtained for each x. Upper and lower bounds were estimated by calculating the 99% and 1% limits. A baseline period was defined between -4.5 and -3 s relative to the time of movement. The lower bound of -4.5 s was chosen to avoid distortions in the density function due to edge effects. The distribution of tones is assumed to be uniform in this early period where action intentions are presumably not yet conscious. T-time was defined as the last point at which the 99% bound exceeded the baseline. Figure 7.4 illustrates the effect of increasing the number of bootstrapping samples.

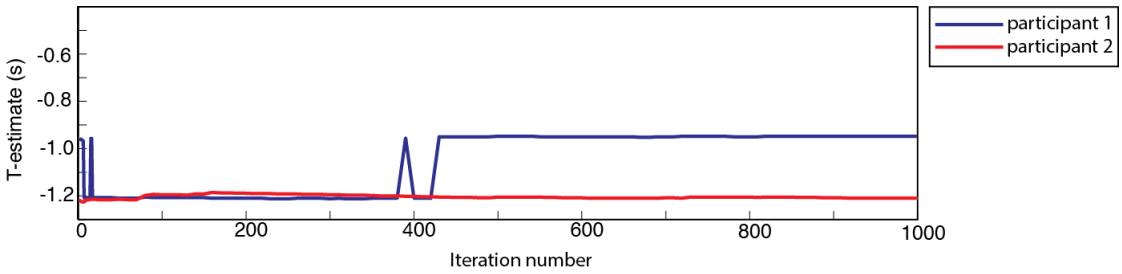


Figure 7.**4** Experimental data from two illustrative participants shows the effect of increasing the number of iterations used in the bootstrapping. In some cases and despite initial plateaus, increasing the number of bootstrap samples can have important effects on the final estimated T-time.

As in Libet's original experiment, participants were encouraged to make movements as soon as they became aware of their intentions to act. It was emphasized that they should pay attention to their intentions, but that they should not introduce deliberate delays between the time of conscious intention and the time of movement. If participants were following instructions adequately, then the estimated T-time should not be excessively early. In the same way, if participants were attending to their intentions, a valid estimated T-time will be earlier than the time of movement.

7. 3 Results

Subjects made movements with an average inter-movement interval of (mean ± SD) 4.810 ± 1.33 s for the instructed blocks and an inter-tone interval of 4.67 ± 1.25 s for the intentional blocks. These values did not differ significantly (t_{15}=0.64, p=0.527).

Participants reported having inhibited relatively few movements per block of 35 tones. Roughly 5-8 out of 35 tones were reported to lead to action inhibition.

Two different methods were used to estimate T-time. First, a parametric method was used, and individual T-times were defined as the value of the p_2 parameter (see equation 7.2), that determines the position of the step of the *erf* function along the time *(x)* axis. Second, a density function was estimated (see methods).

erf fit

The data of 1 participant were discarded because the T-time estimate could not be obtained with the erf method. The data for one representative participant are shown in figure 7.5.

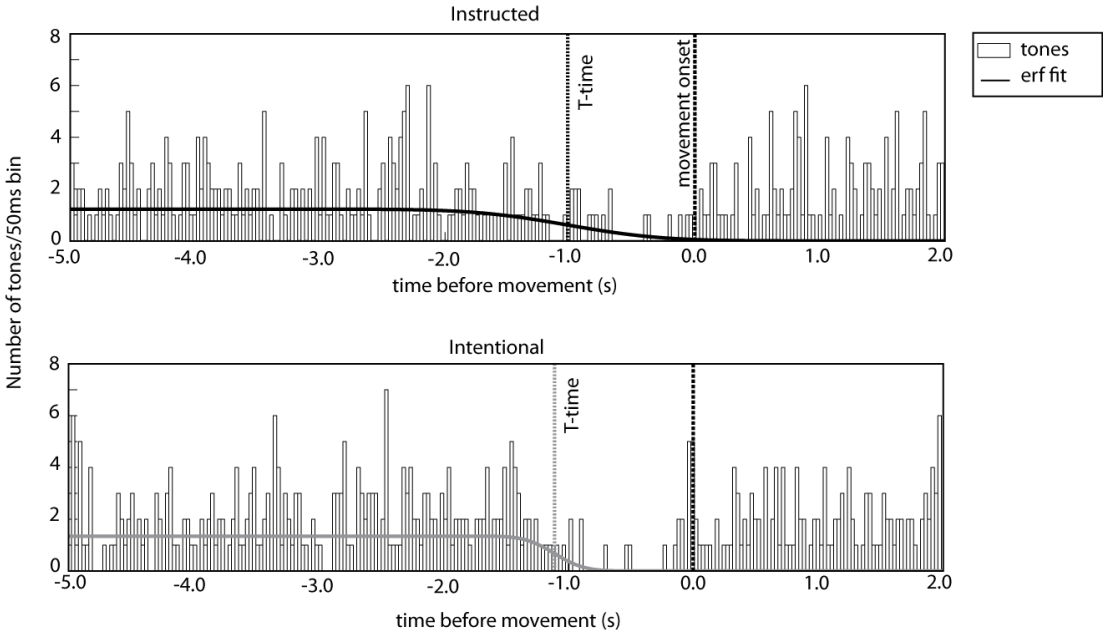


Figure 7.5 Tone frequency distribution and estimated T-time for one representative participant. Solid lines show the fitted *erf* function **Panels A. and B.** show histograms for the instructed and intentional conditions respectively.

Mean (± SD) estimates of T-time were -1.128 ± 0.724 s and -1.095 ± 0.635 s for the instructed and intentional conditions respectively. These values did not differ significantly (t_{14} = -0.26, p = 0.796).

Table 7.1 Individual T-time estimates (in seconds) for the instructed and intentional conditions, obtained with the parametric (*erf*) or nonparametric (density estimation) methods.

	***erf* fit**		**Density estimation**	
Participant	Instructed	Intentional	Instructed	Intentional
1	-1018.25	-1110.12	-1247.5	-1257.5
2	-2000.88	-677.97	-2722.5	-300
3	-596.15	-347.38	-592.5	-350
4	-1406.99	-1128.55	-2535	-2012.5
5	-800.55	-919.28	-320	-527.5
6	-546.48	-606.33	-1347.5	-677.5
7	-900.00	-1008.33	-1147.5	-1192.5
8	-700.32	-1797.70	-2930	-2302.5
9	-521.50	-577.53	-617.5	-480
10	-2887.19	-2797.36	-3127.5	-3572.5
11	0.00	-395.67	-342.5	-332.5
12	-1814.48	-1606.93	-300	-2607.5
13	-1636.77	-1498.78	-300	-300
14	-896.19	-853.22	-1400	-1037.5
15	-1200.00	-1100.00	-542.5	-300

Estimation of density function

T-time was then estimated by a nonparametric method, estimating a density function and bootstrapping the estimated density to obtain a closer fit to the data.

The density estimation method presents two main advantages over the parametric method. First, the results from the parametric method vary with the bin width and bin centres chosen to construct the frequency histogram; and present a bias of order h for bins of size h. In contrast, the density estimation with symmetrical kernels (such as the gaussian kernels used here) does not dependent on the arbitrary bin parameters (Sheather, 2004).

Figure 7.6 shows data from one representative participant. Mean (± SD) estimates of T-time were -1.300 ± 1035 s and -1.150 ± 1.022 s for the instructed and intentional

conditions respectively. These values did not differ significantly (t_{14}=-0.61, p=0.553).

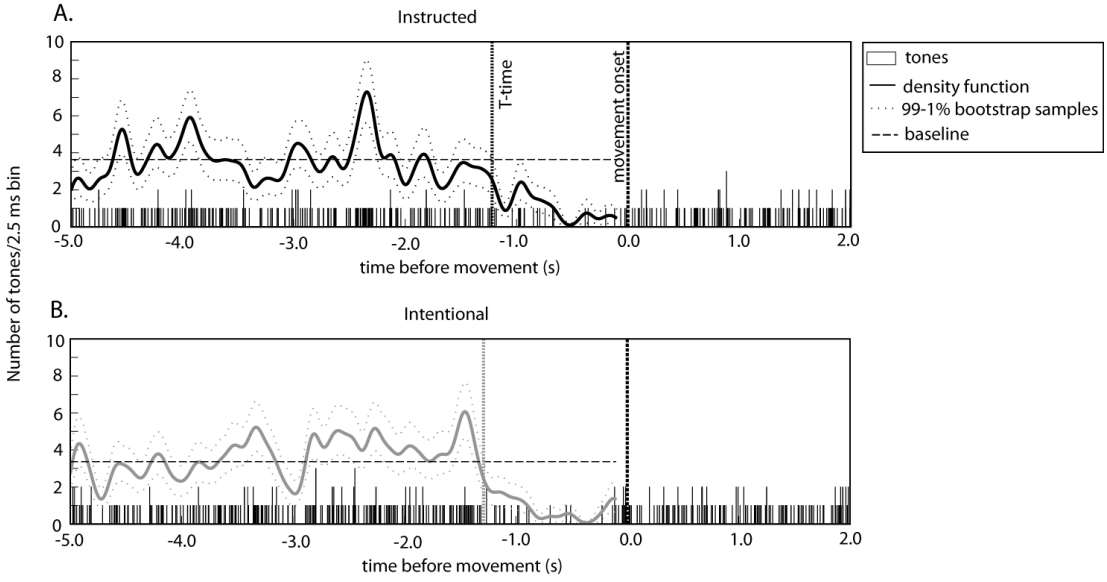


Figure 7.6 Tone frequency distribution and estimated T-time for one representative participant. Solid lines show the estimated density function **Panels A and B** show histograms for the instructed and intentional conditions respectively.

Comparison of the two methods

Matsuhashi and Hallet applied both methods (*erf* fit and density estimation) to estimate T-time. They found that the density estimation method yielded slightly earlier T-time estimates. The 95% confidence intervals of the difference between the two methods was from 8.1% to 0.3% of the average estimated T-time, suggesting that the methods produced consistent results in their case.

In this study, to explore the agreement of the two estimation methods used, the two-way mixed model intraclass correlation coefficient (ICC) was calculated for each condition (McGraw & Wong, 1996). Values of ICC close to 1 indicate good agreement, and values close to 0 indicate poor agreement between measures. In both cases, ICC suggests strong agreement between measures (instructed: ICC = 0.63; intentional: ICC=0.86. As a rule of thumb, ICC higher than 0.6 can be taken as a measure of strong agreement, and ICC values higher than 0.8 suggest very strong

agreement). Figure 7.7 shows the individual data for each condition, and the linear fits to demonstrate correlation.

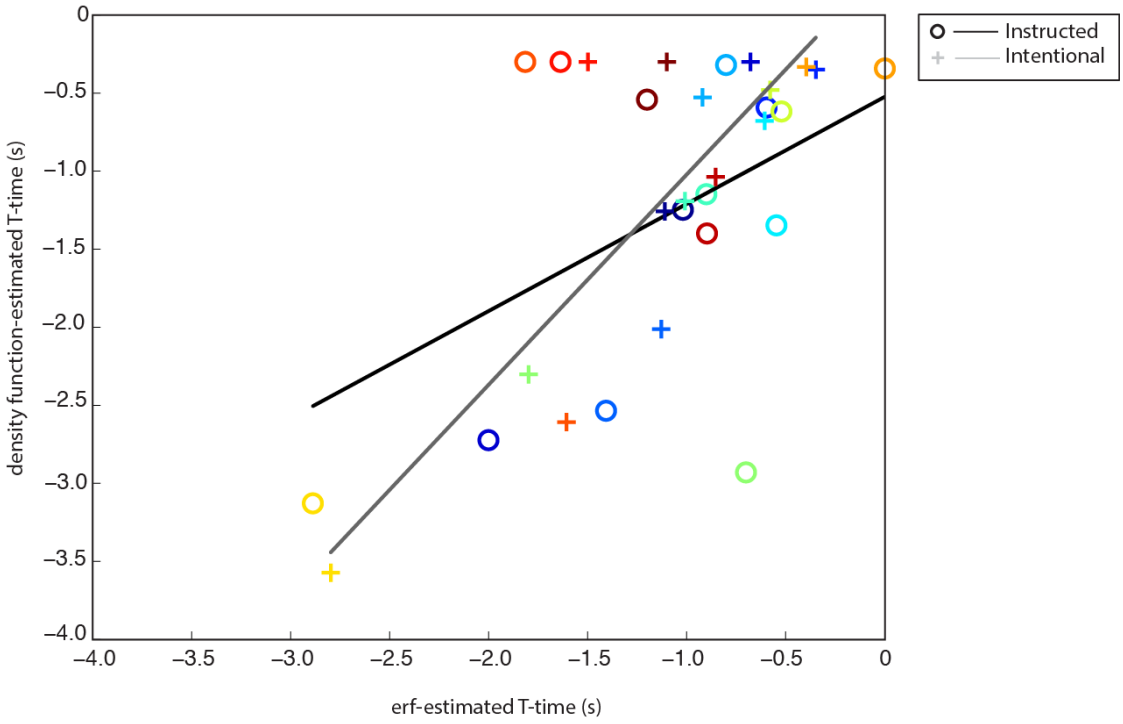


Figure 7.7 Consistency between the two estimation methods. Individual colours represent the pair of the estimated T-times for each participant. Open circles correspond to instructed conditions, crosses correspond to intentional conditions. Trend lines show linear fit for the data of each condition.

7. 4 Discussion

The estimation of the time of conscious intention has traditionally relied on explicit measures that depend heavily on retrospective reports. These measures are problematic because they may suffer from reconstructive biases (Wegner, 2002). Further, retrospective timing judgements may suffer from distortions that are large when compared to relatively subtle effects of interest. For example, this study aimed at examining presumably small differences in the timing of awareness of intention between intentional and instructed actions. Large inaccuracies due to reconstructive errors may greatly decrease the sensitivity of a method in detecting differences between conditions.

A new method has recently been proposed (Matsuhashi & Hallett, 2008) that probes conscious intention online and therefore depends only minimally on retrospective confabulations. The present study aimed at exploring the advantages of the online method, and its potential suitability for detecting subtle differences in the timing of intentions across different conditions. This is the first replication and extension of the online method devised by Matsuhashi and Hallett (2008). This study also appears to be the first sensitivity analysis of a potentially important method.

Replicability of the online method

Importantly, the T-time estimates obtained here are in agreement with those reported by Matsuhashi and Hallett (2008). This shows that the online method is replicable and may contribute to reliable measures of the time of conscious intentions.

However, despite the clear advantages of the online method, a more thorough analysis suggests that it may not be sensitive enough to detect small differences between closely related conditions.

Numerical discrepancies between retrospective and online tasks

There was a numerical discrepancy between the T-time estimate in the online task (~ -1.5 s prior to movement onset) and the typical W-judgement in a Libet-type task (~ -0.2 s prior to movement onset, Libet et al., 1983). Apart from explanations based on differences between attention allocation in the two different tasks, Matsuhashi and Hallett have suggested an interesting possible account. They note that the online task is similar to studies on mind wandering (Smallwood & Schooler, 2006). In their studies, Smallwood and Schooler asked participants to do a repetitive task, and measured the frequency of mind-wandering events using two possible methods of subjective report. First, participants were simply asked to spontaneously inform the experimenter if and when they realized that they were mind wandering. Second, the experimenter periodically probed participants, and explicitly asked them whether they were mind wandering or were focussing on the task. These two methods are interestingly close to the retrospective and online methods for the monitoring of

conscious motor intentions. Whereas in the retrospective task participants need to "wait for the conscious intention to appear", in the online task participants are periodically probed by the auditory tones. Smallwood and Schooler noted that the frequency of mind wandering episodes obtained through spontaneous report was lower than that obtained through direct probing. The notion is that awareness of one's own mental states (mind-wandering in one case, motor intentions in the other) progresses through three states of metacognitive awareness. These are a first stage of no awareness, an intermediate stage of awareness only if one is probed, and a third stage of clear metacognitive awareness. Whereas the W-judgement is related to the transition from the second to the third stages of awareness, T-time estimates in the online task may reflect the transition from no awareness to the intermediate stage.

These interesting differences between the retrospective and online tasks call for richer models of conscious intentions. In particular, mental chronometry methods treat intentions to act as being states that are either present or absent. However, it seems as though this model may be too simple. Instead, intentions may grow in accessibility and clarity as the moment of action initiation approaches.

Limitations of the online method

The online method was evaluated as a possible candidate to detect differences between the intentional and instructed actions. The results show that despite the clear appeal of its potential to avoid certain biases, the task suffers from a number of limitations. These limitations, described below, may make it insensitive to small differences in the timing of conscious intentions between two closely related conditions.

Difficulties of introspecting about intentions to make simple movements

First, index finger extensions may have been excessively simple, and may have not been effective targets for the introspection of conscious motor intentions. More effortful movements requiring more precise force control (Masaki, Takasawa, & Yamazaki, 1998), selectiveness (Kitamura, Shibasaki, & Kondo, 1993) or involving

sequences rather than simple movements (Kitamura, Shibasaki, Takagi, Nabeshima, & Yamaguchi, 1993) have all been shown to elicit larger amplitudes in the late RPs. Replacing single index extensions for other, more effortful movements may result in enhanced differences in amplitude of RPs and in the timing of conscious intentions between intentional and instructed conditions.

Reliance on subjective report

Second, the online task suffers from the same major caveats as the retrospective task, in that it critically relies on subjective report. This is especially problematic because the required introspection about conscious intentions is unusual and effortful. Retrospective tasks directly measure the time of conscious intention along a continuous variable (i.e., position of the clock hand). In contrast, the online task measures binary responses (i.e., if the participant is aware of her intentions at the time of tone, she will inhibit the movement. If she is not aware of her intentions, she will ignore the tone). However, both continuous and binary responses require that participants periodically introspect on their conscious motor intentions.

Lau *et al* (H. Lau, Rogers, & Passingham, 2006b) asked participants to perform a modified Libet task in an fMRI scanner. The authors compared BOLD activity in conditions where participants judged the time of movement (M-judgement) or the time of an external tone. They found that BOLD activity in the CMA correlated negatively with the amount of temporal bias in the M-judgement. Further, the authors found in a separate dataset that when participants were asked to attend to the time of intention (W-judgement), activity in the SMA was negatively correlated with the amount of temporal bias in the W-judgement. The authors argue that these correlations reflect the attentional modulation of the BOLD activity of those brain structures that support action timing (cingulate motor area, CMA) and intention generation (SMA). They therefore argue that introspective methods may be problematic for neurophysiological measurements, because the required introspective judgement needed modulates the brain activity precisely in those areas that are the targets of study. If this argument is true, the method for reporting awareness may interfere with the neurophysiological processes that cause awareness.

The online task does not offer a solution for this problem. In fact, it may even enhance this interference effect, as attention is drawn away from the otherwise external "Libet clock" and focused exclusively on the abstract internal representations of intentions.

Undersampling problem

An examination of the data for individual participants suggests that the number of tones was not enough to produce a smooth frequency function in continuous space. Data from several participants showed abrupt falls from the baseline tone frequency (of mean 1) to 0. This suggests that sampling larger numbers of trials might have produced smoother distributions. More importantly, data from some participants showed isolated tones in the period around -3 s prior to movement. Arguably, these tones were too early to fall within the point of no return (de Jong et al., 1990). The reason why they were immediately followed by a movement (instead of leading to inhibition of movements) may have been simply due to mind-wandering effects. Participants may have been distracted from their intentions, and may have not responded correctly to the tones. Mind-wandering and attention drifts are common in most experimental tasks (Smallwood & Schooler, 2006) and are usually overcome by introducing a large number of trials, i.e., by treating them as random noise . However, due to the low frequency of inhibited movements in the online task, these "incorrect" trials may have an unusually large detrimental effect. These two issues may be solved by greatly increasing the duration of the task, or perhaps by making the task less monotonous and interleaving intentional and instructed movements. Because individual trials cannot be readily identified in the online task, a paradigm that includes shorter blocks may help reduce the number of such trials.

Time of conscious intention is indirectly measured

Finally, unlike retrospective methods, the online method depends on the estimation of a putative underlying function. As such, the final result of the online method depends heavily on the arbitrary choices of the free parameters chosen for the estimation of the function. Importantly, neither estimation method aims at minimizing residual errors. Instead, the fit of the estimated function to the tone distribution is a non-optimal one. The parametric estimation of the *erf* function relies on a precise fit of the time at which the frequency distribution decays. However it does not attempt to minimize the large residual fitting errors. These result from approximating a constant function to a variable rate of tones. In the same way, the density estimation method requires an arbitrarily chosen bandwidth, that has a critical effect on the final result (Sheather, 2004) (see figure 7.3). Several methods have been suggested to optimize this free parameter and avoid arbitrary choices e.g., (Shimazaki & Shinomoto, 2010). However, the desired density function is not a precise and optimal estimation of the data. Instead, and critically, it depends on approximating a steady baseline to a period of non-steady tone frequency. Thus, by definition the density estimation must be non-optimal and optimization algorithms are not applicable here. Exactly how non-optimal the fit should be is not clear, and not quantifiable.

Dependence on the task parameters

Importantly, these results highlight the possibility that the measure of subjective experience may not be independent from the method used to address it. In particular, the frequency of tones may influence the reported time of conscious intentions. The time between the emergence of a conscious intention to move and the movement itself may be short (under 1 s, as reported by Libet *et al*). The tones in this task were presented with a relatively long inter-tone interval (2-5 s). Therefore, tones were unlikely to fall precisely within the narrow time window during which participants were aware of their motor intentions. Consequently, to comply with the task instructions, participants may have deliberately introduced delays between the time of intention and the time of movement, in order to allow tones the possibility to interrupt their prepared movements. Because these presumed pauses are not

accessible to objective analyses, there is a risk that conscious intention is brought backwards in time by an arbitrary amount. The instructions may have been an important part of this task. Indeed, other work in this thesis (see chapter 4) confirms that participants can transiently inhibit action in this way. If this speculation is true, then presumably decreasing the inter-tone interval will lead to shorter T-time estimates. In the same way, instruction that stress the need for interruptions of prepared movements may lead to longer T-time estimates.

7. 5 Conclusion

The online method suggested by Matsuhashi and Hallett (2008) was applied here, and their results were replicated. The estimated values for the mean time of thought (T-time) in this study were similar to those reported by Matsuhashi and Hallett originally.

However, a closer examination of the online method revealed some methodological difficulties that did not easily emerge from Matsushashi and Hallett's original proposal. The online task has been shown to introduce a large set of potential inaccuracies, and cannot provide a final answer for the question.

Therefore, the question of whether awareness of action is earlier for intentional actions, than for instructed actions, remains unclear. A definite answer will require the development of new paradigms that work around the caveats that must always be issued when conducting experiments with the currently available methods.

Chapter 8 Brain correlates of the subjective feeling of freedom of choice

A classical experimental design contrasts instructed and intentional actions, based on the objective definition of the extent to which actions are constrained by an external stimulus. Instead, in the experiment reported here, intentional and instructed behaviour in a numerical stem completion task were defined subjectively, based purely on participants' introspective reports of how free their responses felt to them. An analysis of the blood-oxygen level-dependent signal increases associated with intentional action as defined in the classical, objective way, and as defined in this new, subjective way, revealed striking differences. The neural correlates of feeling free did not overlap with the neural correlates of objectively being free, in the classical conception of being stimulus-independent.

8. 1 Introduction

Debates over whether humans have the capacity to make free choices have been ongoing for countless years (Libet, 1999; S. Thomas, 1894; Wegner, 2002). In contrast, there is wider consensus about the existence of a *subjective feeling* of acting freely (Sarkissian et al., 2010). According to folk psychology, and consensus of experience, people generally have the impression that their internal conscious decisions drive their behaviour. In other words, people's decisions and actions are not simple reflections of the immediate environment, but rather expressions of an "agentic self" (Kane, 2005; Schüür & Haggard, 2011).

According to one view, these subjective experiences are illusory (Wegner, 2002), and the "conscious will" is merely a retrospective inference, rather than a direct readout of brain activity associated with action selection or action generation. Supporters of this view often draw on behavioural studies that reported illusions of will and agency (Wegner & Wheatley, 1999), and on neurophysiological measurements (Libet et al., 1983; Soon et al., 2008) showing that neural events associated with free decisions precede the reported onset of the awareness of intention. These findings suggest that conscious intention cannot be a causal factor for free decisions. Interestingly, despite the recent scientific support for 'free will illusionism', few studies have investigated where in the brain this alleged illusion arises.

On the other hand, several human neuroimaging studies have shown reliable neural correlates of free choice. The contrast between free and instructed movement choices has been consistently associated with increased BOLD signal in the SMA and preSMA, the rostral cingulate zone (RCZ) and the dorsolateral prefrontal cortex (DLPFC) (Cunnington, Windischberger, Robinson, & Moser, 2006; Lau, Rogers, Haggard, & Passingham, 2004; Lau, Rogers, & Passingham, 2006), (see Krieghoff, Waszak, Prinz, & Brass, 2011) for a review). In particular, Müller *et al*. (Mueller et al., 2007) have suggested that RCZ is mainly involved in selecting the "what" component, in the context of a given task (Desmet, Fias, Hartstra, & Brass, 2011); while preSMA is associated with selecting the "when" component of actions.

These neuroimaging studies generally used an objective manipulation in the experimental design factors, defining instructed and free choice in terms of information that either respectively is or is not provided by external cues. This definition bypasses the subjective experience of free choice. However, understanding the neural basis of the common feeling of acting freely is important. On the one hand, understanding the mechanism underlying illusions has long been a productive approach in psychological research, and so would be a natural research question for free will illusionism. On the other hand, if 'free choice' is not an illusion but a distinctive psychological form of decision-making, then the neural bases of the subjective feeling of free choice may be relevant to understanding how and where such choices occur in the brain.

The scientific tradition reviewed in the introduction (see section 1.2.2.2) operationalizes voluntary action based on objective criteria. This operationalization is implicitly assumed to capture the subjective feeling of acting intentionally (Schüür & Haggard, 2011). However, this important implicit assumption has never been appropriately validated. If the assumption is correct, and there is a correspondence between objective and subjective accounts of free action, then the neural correlates of free and instructed choices defined objectively should match the neural correlates of choices that feel more versus less free. An experimental task was therefore devised, in which actions were defined as instructed or free either on the basis of an objective definition, or on the basis of subjective experience. Then, the brain correlates associated with free choices under each of the two possible definitions were investigated.

Here the classic distinction between instructed and intentional choice was used and extended to investigate the neural correlates of subjective voluntariness. Importantly, it was also considered that external guidance can come in varying degrees; so that the instructed/intentional distinction is not a simple dichotomy between two exclusive categories, but rather represents two extremes of a continuum (Nachev, 2010). On this view, generating an action can involve both intentional and instructed factors. For example, responses to external stimuli clearly depend on longer-term goals that are represented internally. For the purposes of this study, intentional choice was considered as a *graded* measure of how independent an action is from an external instruction. When an action is strongly determined by an external

instruction, it will be "less free" than when it is not. It may then be asked whether people subjectively experience degrees of voluntariness underlying individual action decisions, and whether this graded experience originates from graded levels of activation in particular brain areas.

Here, the classic task of random number generation was adapted from previous studies of free action choices (Jahanshahi et al., 2000). This task was chosen because it allows both an objective, graded continuum of integration of stimulus information, and also a graded continuum of subjective experience of voluntariness regarding action choice. Random number generation tasks have been used before in relation to voluntary action, but for rather different reasons from the ones here. In particular, many intentional selection studies involve asking participants to produce balanced numbers of responses, while avoiding obvious patterns such as alternation. These tasks have been interpreted as covertly asking participants to generate apparently random response sequences (Roepstorff & Frith, 2004). For example, human positron emission tomography (PET) studies during random number generation tasks have suggested a critical involvement of the left dorsolateral prefrontal cortex (DLPFC), the anterior cingulate cortex, the bilateral superior parietal cortex, and the right inferior frontal cortex (Daniels, Witt, Wolff, Jansen, & Deuschl, 2003; Jahanshahi et al., 1995). These areas partially overlap with those identified with free selection tasks (Jahanshahi et al., 1995). Here, a modified random generation task offered a convenient vehicle to allow participants to experience and report a graded sense of stimulus-independence or freedom of action.

Crucially, two separate analyses were performed. First, the extreme situations of intentional and instructed action were compared, as operationalized by the classical objective paradigms. Second, situations in which actions subjectively *felt* more free or *felt* more constrained, based on self-report, were studied. In this way, this study aimed to establish the relationship between the neural correlates of intentional actions as traditionally defined, and the neural correlates of the subjective experience of choosing freely.

In the crucial condition for studying graded voluntariness, a number sequence was presented and participants were asked to complete the sequence with a number that would make the sequence "look random". The method that participants used to achieve this presumably varied from person to person, depending on their subjective

concept of a random sequence. However, the actual interpretation of "random appearance" was not the central interest here: trials were classified based on subjective report of how free the choice of action was felt to be. In contrast, the focused was on participants' subjective feeling of voluntariness associated with whatever choices they had in fact made, and not on the mechanisms that caused them to make those choices.

8. 2 Methods

Participants

Twenty-three healthy participants took part in the study (5 female; mean age 22±2 years). All participants gave written informed consent. Procedures were approved by the local ethical committee, and were in accordance with the Declaration of Helsinki. All participants had normal or corrected-to-normal vision. No participant had a history of neurological, major medical, or psychiatric disorder.

Stimuli and procedure

Each participant made manual actions to choose numbers on a screen using a trackball in experimental trials within three different contexts in a random number generation task.

Each trial proceeded as follows (see figure 8.1). A white fixation cross was displayed on a black background. The duration of the fixation cross was sampled from a pseudologarithmic distribution, and ranged from 2 to 14 s. The fixation cross served as a variable period constituting an implicit baseline BOLD measure. Immediately after, either a sequence of four numbers (from 1 to 4) or four X's appeared on the screen, above a 2x2 response grid. The position of each number in the grid was randomly assigned and changed in every trial. Participants held an MRI-compatible trackball on their lap. The mouse cursor was displayed on the screen using a red "+" sign, initially positioned on the centre of the response grid. Participants were instructed to select a number from the response grid by moving the mouse cursor to the chosen number and clicking on it. The choice of number was based in the

number stem presented (see below). Once the number was selected, it was displayed next to the stem for 0.8 s.

In two of the contexts the question "How free was your choice?" ("Hoe VRIJ voelde je keuze aan?" in Dutch) was displayed above a visual analogue scale (VAS). The VAS had 10 subdivisions, and its extremes were labelled "Very free" and "Not free" ("HEEL vrij" and "NIET vrij"). The left-right orientation of the VAS labels was counterbalanced across participants, but kept constant for each participant throughout the practice and experimental sessions, to avoid confusion. Participants were asked to indicate how free they felt their choice had been by clicking with the cursor on the appropriate position. Participants were reminded that they could use the whole range of the VAS.

The maximum response time for the number selection and the voluntariness rating was 5 s. If participants had given no response after this time had elapsed, a message appeared and the next trial started.

First, conditions of intentional choice versus instructed choice were compared in a *classical* context (see below), operationally defined in the same way as the classical literature on free/instructed choices. Second, in an *objective* context, participants performed the free and instructed conditions again, but each trial was followed by a rating of how free they felt their immediately-preceding action choice had been. Finally, in the crucial *subjective context*, participants chose which of several actions to make following presentation of a suggestive stem stimulus, and rated on a continuous scale the extent to which their action choice had been free or not free, with respect to the information given in the suggestive stimulus.

The subjective context necessarily involved a subjective rating, whilst the classical context did not, making it difficult to compare these conditions directly. The objective context was included to get round this problem, by providing a condition that was informationally equivalent to the classical context, but also included the element of introspective report. In this way, BOLD activity associated with the objective context could be contrasted with BOLD activity in the subjective context, because both contexts included both an action selection and a judgement event in each trial.

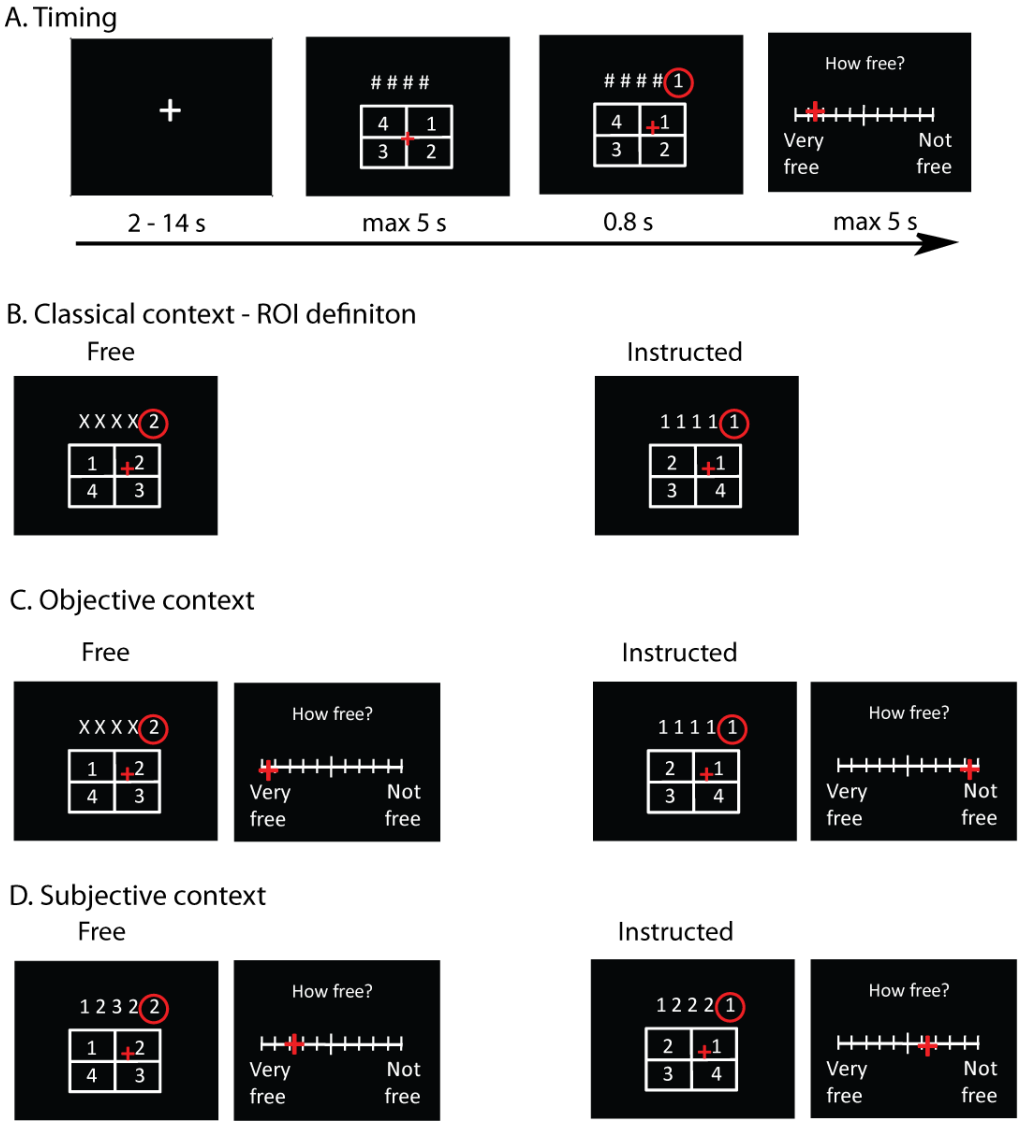


Figure 8.1: Experimental paradigm. A. General timing of events. Four numbers or four Xs were displayed on the screen (in the figure, # represents either a number or an X. The symbol # was never actually displayed in the experiment). Participants chose a number by clicking on it with a trackball. Participants were then asked to click on a visual analogue scale to rate how free their choice had been. **B.** The classical context was used to identify regions of interest associated with the contrast free > instructed in independent data. In the objective context (**C**) trials were defined as free or instructed *a priori*. In objective free trials, a series of four X's was displayed and participants were free to choose any number from the response grid. In objective instructed trials, participants saw a sequence of four identical numbers ("1 1 1 1" in this example) and participants were asked to complete the sequence by clicking on the same number that was displayed. Participants then indicated their feeling of voluntariness. In the subjective context (**D**) four numbers were displayed, and participants were asked to complete the sequence with the fifth number "in order to keep the sequence looking random" (see Methods for details). Trials were then classified *a posteriori* according to a median split of each participant's subjective reports.

Classical context

The *classical* context was designed to identify patterns of BOLD signal associated with free actions as classically defined (Richard E Passingham, Bengtsson, & Lau, 2010). In free trials, the presented sequence was always "X X X X". In these trials, participants were free to choose any of the four numbers (1, 2, 3 or 4) displayed on screen as their action. In instructed trials, the sequence contained one single number repeated four times (e.g., "1 1 1 1"). In these trials, participants were instructed to choose the number that was displayed on the screen (i.e., 1). The *classical* context was effectively used as a localizer, to define regions of interest (ROIs). The BOLD activity for the Objective and Subjective contexts (see below) was analyzed for ROIs identified independently from the data from the classical context.

Objective context

The main aim of this experiment was to compare the brain correlates of the extreme free and instructed conditions with those of a subjective context in which participants could themselves report how free their choice had been by means of a VAS (see below). In order to make the two contexts comparable, a VAS rating was included in the objective context. In this way, any BOLD activity differences between the objective and the subjective contexts could not be attributed to the mere presence of the VAS.

Subjective context

The *subjective* context aimed to provide participants with graded experiences of more or less free choice. Unlike in the other two contexts, the numerical stimulus presented in each trial contained a pseudorandomized sequence of numbers. Participants were instructed to use this sequence as a suggestive guide for their response. They were asked to choose a number that would make the stimulus sequence "look random". It was assumed that folk knowledge would guide participants in their choices (Nickerson, 2002). Each participant might have felt highly constrained by some preceding sequences, and very unconstrained by other sequences. It was assumed that participants would report feeling less and more free in those trials respectively. A random sequence generation task was used as a means

to provide participants with a graded and reportable experience of voluntariness, in the sense of freedom from constraint. This could then be used this to investigate the brain activity associated with the experience of voluntariness.

The numerical stimulus therefore served to prompt the next action choice to some extent. The extent to which it did so was assumed to influence feelings of subjective freedom of choice. Any of a number of rules could relate the numerical stimulus to the chosen action. Participants were asked to make the stimulus "look random", and to report their subjective feeling of freedom of choice. The precise *form* of the completion rule used may vary across participants and was irrelevant to the purposes here. Instead, the main interest was the *extent* of constraint provided by the numerical stimulus and the self-selected rule. It was assumed that subjective reports of freedom of choice indicated this extent. In this way, voluntariness was not directly manipulated experimentally. Instead, introspection was taken as a reliable method to report the subjective experience. The extent of free choice was thus a dependent variable, in contrast to its normal status as an independent variable in the classical action selection literature.

The sequences for the subjective context were generated by a pseudorandomized procedure. The measure of "stimulus space" was defined for each sequence as the number of different numbers present in the stem (irrespective of position). Thus, "3 3 3 3"; "2 2 1 2"; "3 1 4 4" and "2 3 4 1" are sequences with stimulus spaces of 1, 2, 3 and 4 respectively. The total of 160 trials in the subjective context were divided into 4 blocks of 40 trials. Pilot results suggested that participants' ratings of subjective freedom of choice were related to the size of stimulus space. Therefore, each block contained trials with all possible stimulus spaces in equal proportions. The sequences examplars for the subjective context were the same for all participants, to allow for potential comparisons across participants. Their order of appearance was randomized across blocks and trials.

Additionally in the subjective context, a memory question was displayed every 10 trials. A four-number sequence was presented and participants were asked if that sequence had been presented in the preceding 10 trials, with a maximum response time of 5 s. This memory question was aimed at encouraging participants to pay attention to the number stimuli presented on every trial. The responses were not analyzed.

Each participant performed two consecutive blocks of 40 trials in each of the classical and objective contexts. They also performed 4 consecutive blocks of 40 trials in the subjective context, which formed the key focus of the study. The order of the contexts was randomized across participants.

Before scanning, participants were trained with at least one practice block for each context, always in the same order: classical; objective, subjective. Training continued until participants felt comfortable with the task. The experiment in the scanner lasted approximately 70 min.

After scanning, participants completed five personality questionnaires addressing feelings of control (Rotter, 1967) belief in free will (the free will and determinism scale, (Rakos, Laurene, Skala, & Slane, 2008); the social desirability scale (Crowne & Marlowe, 1960); and two self-control questionnaires (Rosenbaum, 1980; Tangney, Baumeister, & Boone, 2004). In addition participants completed a semi-structured questionnaire about the strategies they had adopted in the completion of the random number sequences.

fMRI data acquisition

Participants were positioned head first and supine in the magnet bore. Images were collected with a 3T Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany), using an 8-channel radiofrequency head coil. First, 176 high-resolution anatomical images were acquired using a T1-weighted 3D MPRAGE sequence [TR = 2500 ms, TE = 2.58 ms, image matrix = 256 × 256, FOV = 220 mm, flip angle = 7°, slice thickness = 0.90 mm, voxel size = 0.9 mm × 0.86 mm × 0.86 mm (resized to 1 mm × 1 mm ×1 mm)]. Whole brain functional images were collected using a T2*-weighted EPI sequence, sensitive to BOLD contrast (TR = 2000 ms, TE = 35 ms, image matrix = 64 × 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 17%, voxel size 3.5 mm × 3.5 mm × 3 mm, 30 axial slices). A varying number of images were acquired per run due to the self-paced initiation of trials.

Data processing and analysis

Trials with reaction times (RTs) (for either the number choice or the voluntariness rating) shorter than 0.2 s or longer than 5 s were discarded from the analysis. Instructed trials with incorrect responses were also discarded.

The fMRI data were analyzed with statistical parametric mapping, using the SPM8 software (Wellcome Trust Centre for Neuroimaging, University College London, London, UK). The first four scans of all EPI series were excluded from the analysis to minimize T1 relaxation artefacts. A mean image for all scan volumes was created, to which individual volumes were spatially realigned by rigid body transformation. The high resolution structural image was coregistered with the mean image of the EPI series. The structural image was normalized to the Montreal Neurological Institute template. The normalization parameters were then applied to the EPI images to ensure an anatomically informed normalization. A commonly applied filter of 8 mm FWHM (full-width at half maximum) was used. The time series data at each voxel were processed using a high-pass filter with a cut-off of 128 s to remove low-frequency drifts. The subject-level statistical analyses were performed using the general linear model. The events were defined as the onset time of the stem and response grid. Movement times were also included as parametric regressors to account for variance associated with simple motor activations All resulting vectors were convolved with the canonical haemodynamic response function (HRF) and its temporal derivative to form the main regressors in the design matrix (the regression model). Realignment parameters in all 6 dimensions were also entered in the model to account for variance associated with head motion. The statistical parameter estimates were computed separately for each voxel for all columns in the design matrix. Contrast images were constructed from each individual to compare the relevant parameter estimates for the regressors containing the canonical HRF. The group-level random effects analysis was then performed. The resulting maps were thresholded with $p<0.001$ and cluster-size corrected by means of Monte Carlo simulation. Accordingly significant effects were reported when the volume of the cluster was greater than the Monte Carlo simulation determined minimum cluster size volume (25 voxels), above which the probability of type I error was below 0.05 (Cox, 1996).

8. 3 Results

Behavioural results

Trials from the classical and objective contexts were classified into "free" and "instructed" according to the stimulus presented in each trial: e.g., "1 1 1 1", "2 2 2 2" etc. were classified as instructed, and "X X X X" were classified as free. Instead, trials from the subjective context were classified into "feels instructed" and "feels free" by means of a median split on the distribution of each participants' subjective reports. The median for each participant was calculated on the basis of all the valid trials (> 0.2 s and < 5 s of reaction time -RT-) across all four subjective context blocks.

Thus, trials in the classical and objective contexts were classified *a priori*; whilst trials in the subjective context were classified *a posteriori*, on the basis of the participants' subjective reports. As a validation of the design and analysis, the overall distribution was examined (for all participants) of the ratings of voluntariness for the objective free and instructed conditions separately. Overall, the proportion of free trials ("X X X X" displayed) subjectively classified as instructed was 4.1%, and the proportion of instructed trials (e.g. "2 2 2 2" displayed) subjectively classified as free was 4.3%. This shows that an *a posteriori* classification criterion (according to the voluntariness ratings) closely matched the *a priori* classification criterion (according to the information presented in the stimulus). The free and instructed conditions were therefore comparable across the objective and subjective contexts.

To evaluate whether there were any differences in behaviour across conditions, mean RTs were obtained for free and instructed conditions (see figure 8.2). There were no differences between the RTs for free and instructed trials in the objective context (t_{22}=0.68, p=0.503). However, in the subjective context, instructed trials had longer RTs than free trials (t_{22}=4.93, p<0.001). Subjective instructed conditions imposed, by definition, more restrictions on the responses available to the participant than the subjective free choice condition. These restrictions will have followed from whatever rules the participant used to generate the response from the numbers presented in the stem sequence. Applying these rules, and defining the available set

of response alternatives, may have been a time-demanding process that led to increased RTs in the instructed conditions.

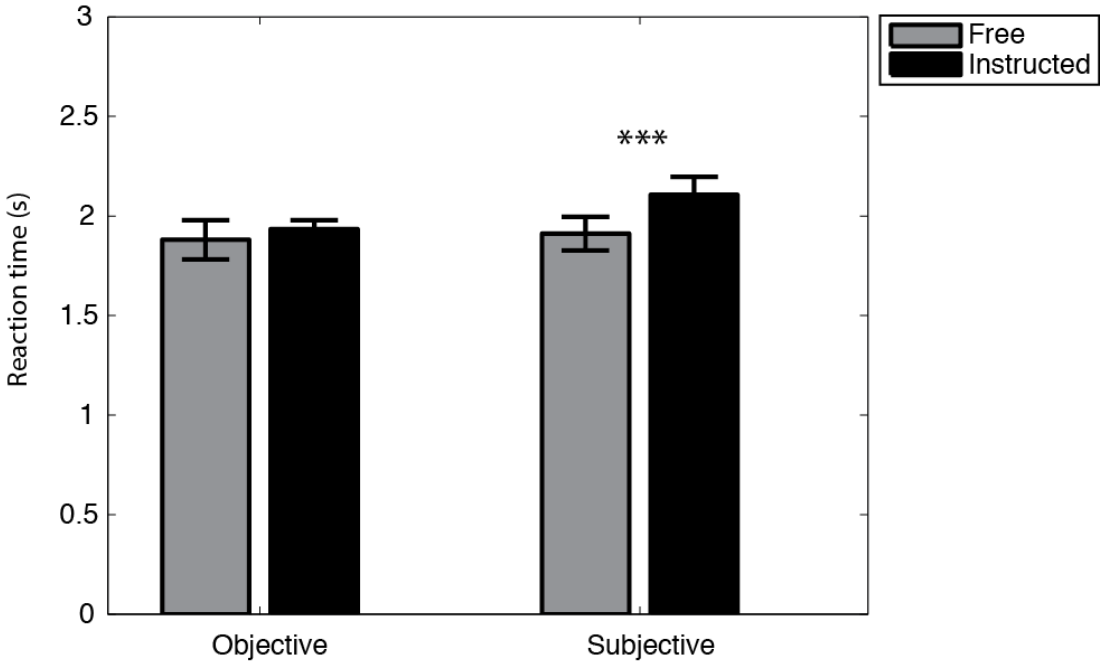


Figure 8.2: Mean reaction times across participants for all free and instructed trials in the objective and subjective conditions. Subjective instructed conditions were associated with longer reaction times than subjective free conditions ($p < 0.001$, see text).

Subjective context – behavioural analysis

As an initial approach to the behavioural analysis of the chosen numbers, the possibility that participants were following an exclusion rule for the stem completion task was examined. Crucially, the dependence between voluntariness rating and exclusion behavior was examined. This analysis provided insight on the factors that made some trials feel more or less voluntary. Exclusion trials were defined as those in which the number chosen in response to the stem was not included in the stimulus space. In contrast, inclusion trials were those in which the number chosen was included in the stimulus space. Trials with no repeated numbers were necessarily inclusion trials, because any number choice would be a repetition of a number already included in the stem. In other words, the dependent variable was not meaningful if all possible responses were already present in the stimulus space.

Therefore, this analysis could only be computed for trials where the stimulus space included at least one number repetition. Results showed that participants predominantly excluded the numbers present in the stimulus space. The exclusion ratio was then computed as the proportion of exclusion trials to the total number of valid trials. The mean exclusion ratio was 0.63±0.15 (± SD), and was significantly different from 0.5, which would have indicated no preference for exclusion behaviour (t_{22} = 4.06, $p<0.001$). However, the voluntariness ratings were not strongly related to a simple factor of exclusion *vs*. inclusion. The voluntariness ratings for exclusion trials and inclusion trials were 4.82±1.54 and 4.08±1.6 respectively, and they were not significantly different (t_{22}=1.7, p=0.1).

To further analyze which factors may have influenced participants' feelings of voluntariness, the number of digits in the presented sequence (i.e., the "stimulus space") was correlated with the degree of perceived voluntariness (see figure 8.3).. A trend analysis revealed a significant and positive linear relationship $F_{1,22}$=18.16, $p<0.001$) between feeling of voluntariness and stimulus space size. A stem sequence with only one number represented (e.g., "3 3 3 3") has a stimulus space of 1. These sequences with small stimulus spaces were associated with responses having the lowest voluntariness ratings.

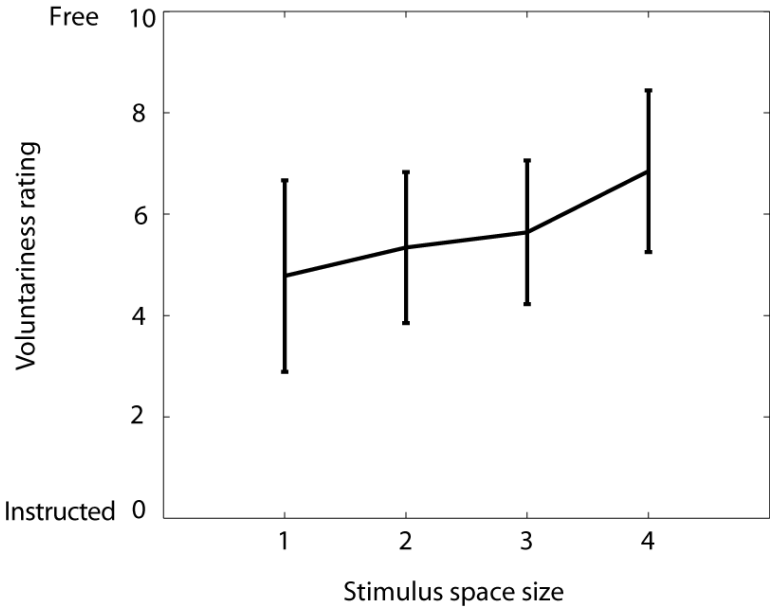


Figure 8.3: Average across participants of feelings of voluntariness as a function of the size of the stimulus space. Stimulus space is defined as the number of different digits present in the sequence shown. The sequences "3 3 3 3"; "2 2 1 2"; "3 1 4 4" and "2 3 4 1" are examples of sequences with stimulus spaces of 1, 2, 3 and 4 respectively.

fMRI results

Classical context, free vs. instructed

First, the free and instructed conditions were contrasted. This contrast served effectively as a functional localizer, based on the objective, classical definition of free and instructed action. The brain areas identified were used to define regions of interest (ROIs) in which BOLD activity for free and instructed trials was analyzed for the independently obtained in the objective and subjective contexts.

The contrast free > instructed in the classical context revealed increased BOLD signal in RCZ/SMA, bilateral inferior parietal sulcus (IPL) left dorsolateral prefrontal cortex (DLPFC) and left premotor cortex (PMC) (see table 8.1). These findings are consistent with existent report of free action > instructed action contrasts (Cunnington et al., 2006; Lau et al., 2004; Lau et al., 2006).

Table 8.1: Results of whole brain analysis in the classical context (free > instructed). The RCZ/SMA cluster extends to both rostral cingulate zone and supplementary motor area and is ambiguously identified by different toolboxes. RCZ, rostral-cingulate zone; SMA, supplementary motor area; IPL, inferior parietal lobe; DLPFC, dorsolateral prefrontal cortex; PC, precuneus.

Area	Peak coordinates (MNI space)			Peak z-score	Cluster size (number of voxels)
	x	y	z		
Right IPL	35	-63	42	5.42	206
Left IPL	-28	-70	46	4.88	217
RCZ/SMA	-4	18	49	4.82	286
Left DLPFC	-35	49	7	4.70	113
Precuneus	11	-70	53	4.62	85
Left PM	-42	4	39	4.20	112

ROI analysis in objective and subjective contexts

The ROIs identified in the free > instructed contrasts in the classical context were tested in independent data from the two other contexts; namely subjective and objective contexts.

The difference between free and instructed conditions in the subjective context was not equivalent to the difference between free and instructed conditions in the objective context, for two main reasons. First, free and instructed conditions in the subjective case were defined by a median split of a feeling of voluntariness that varied along a continuum. In contrast, in the objective context, they were defined as two categorically different situations. Therefore the subjective experiences of voluntariness corresponding to the subjective free and instructed conditions were not expected to be categorically different, in contrast to the objective context. In addition, added "noise" due to errors in subjective report could reduce the strength of the contrast between instructed and intentional conditions in the subjective context, relative to the objective context. For these reasons, a factorial analysis was not

appropriate. Instead, the percent signal change was compared between free and instructed conditions within each context.

This analysis was done within each of the ROIs identified by the free > instructed contrast in the classical context (see Table 8.1). The only difference between the classical context, used to identify the ROIs, and the objective context is that the latter also included a VAS judgement. Thus, unsurprisingly, all six ROIs analyzed showed increased levels of BOLD activity in the free condition as compared to the instructed conditions in the objective context.

Strikingly however, when this analysis was repeated in the subjective context, the pattern of BOLD activity did not match the one in the classical context (see figure 8.4 and table 8.2). Indeed, in five of the six ROIs identified, the difference in BOLD activity between free and instructed conditions was reversed in the subjective context compared to the objective context. In two ROIs (right IPL and left PM) this reversed pattern was statistically significant. In these areas, the BOLD activity was statistically *lower* for actions that felt free as compared to actions that felt less free, but was nevertheless statistically higher for actions that were objectively free, as compared to instructed.

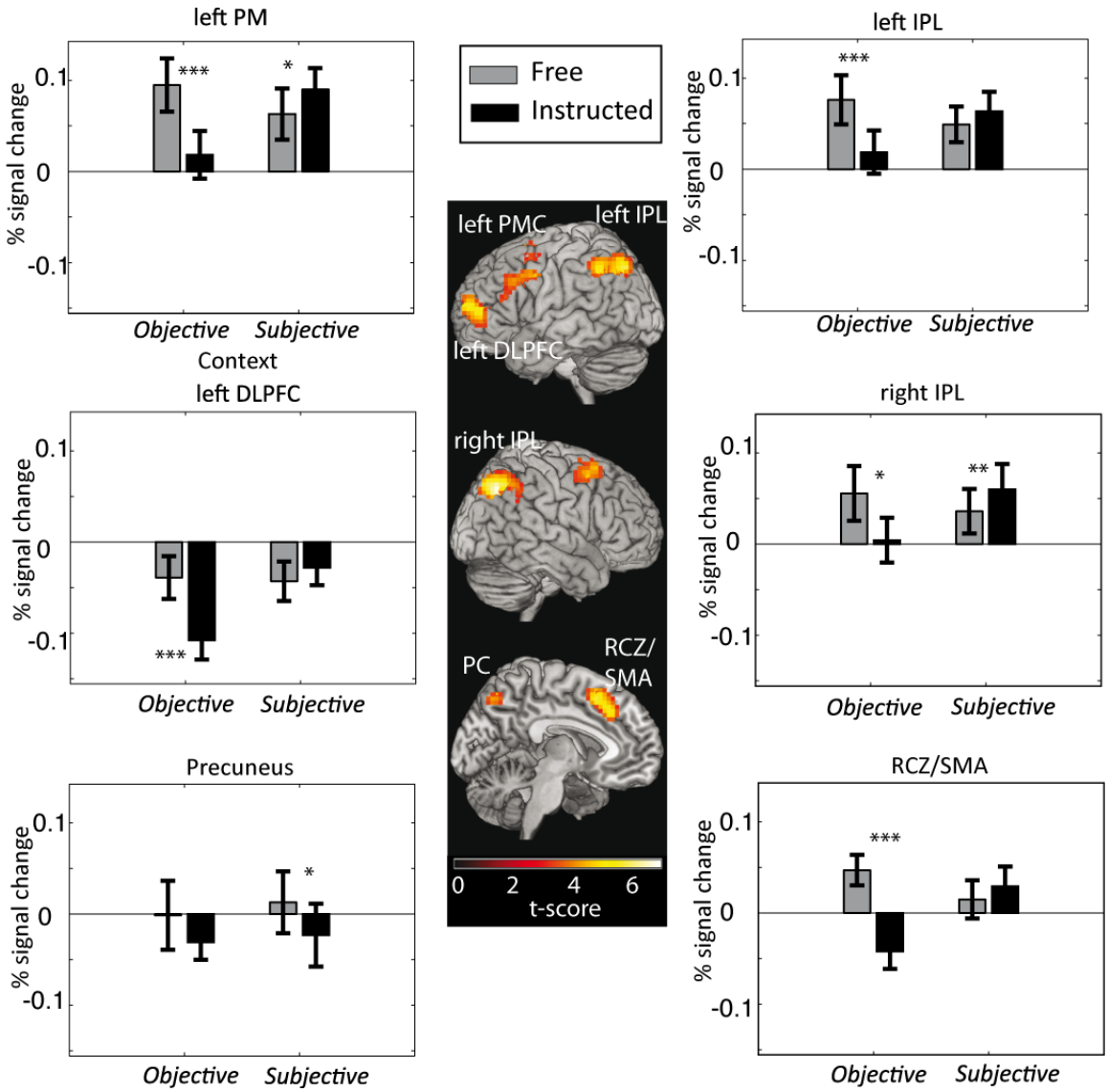


Figure 8.4 Effect of subjective feeling of free action in brain areas defined by objective freedom of action. Results of the ROI analysis for RCZ/SMA, right and left IPL, left DLPFC and left PMC. ROIs were defined on the basis of the free > instructed contrast in the classical context. Percent signal change from data from two independent datasets were then determined in those ROIs. BOLD activations from the objective context correspond to that found in the classical context. On the contrary, BOLD signal from the subjective context follows an inverse pattern in all but one (the precuneus) of the ROIs identified. Error bars show standard error of the mean. RCZ, rostral cingulate zone; SMA, supplementary motor area; IPL, inferior parietal lobe; DLPFC, dorsolateral prefrontal cortex; PC, precuneus.

Table 8.2: Results of statistical tests for each region of interest, as identified by the free > instructed localizer, derived from the independent data in the classical context.

Region of interest (peak MNI coordinates)	**Pairwise comparisons**			
	Free - instructed in objective context		**Free - instructed in subjective context**	
	t_{22}	**p**	**t_{22}**	**p**
RCZ/SMA (-4 18 49)	7.13	<0.001	-1.33	0.19
lDLPFC (-35 49 7)	4.54	<0.001	-1.23	0.23
lIPL (-28 -70 46)	4.83	<0.001	-1.02	0.32
rIPL(35 -63 42)	3.31	0.003	-2.05	0.05
lPM(-42 4 39)	3.80	<0.001	-2.07	0.05
Precuneus (11 -70 53)	1.06	0.302	2.41	0.02

Subjective context – whole brain analysis

BOLD activity associated with voluntary choice as identified by subjective report did not match with that identified by a classical contrast between free and instructed actions. Specifically, in those areas identified by classical free > instructed choice, there was no evidence for stronger activity when participants felt subjectively more free compared to when they felt subjectively less free. In this sense, objectively and subjectively defined free choice did not overlap. However, no direct inferences about interaction effects can be made by the simple combination of two independent t- tests (Nieuwenhuis, Forstmann, & Wagenmakers, 2011). Therefore, to further examine the mechanisms associated with the feeling of voluntariness, and their relationship to the objective operationalization of free action, a whole-brain analysis was done for the contrast free > instructed (median split) in the subjective context. This analysis controlled for effects of stimulus space and RT by including them as regressors in the first-level model. Only one area, in the medial postcentral region, showed increased BOLD signal for this contrast after whole-brain correction by means of a

Monte Carlo simulation (see figure 8.5). BOLD signal showed two peaks of activity in this region, in MNI coordinates (x = 4 y = -21 z = 49) and (x = 0, y = -28, z = 53. This pattern of BOLD activation did not share any commonalities with the pattern found for the contrast free > instructed in the objective context, confirming the results from the ROI analysis

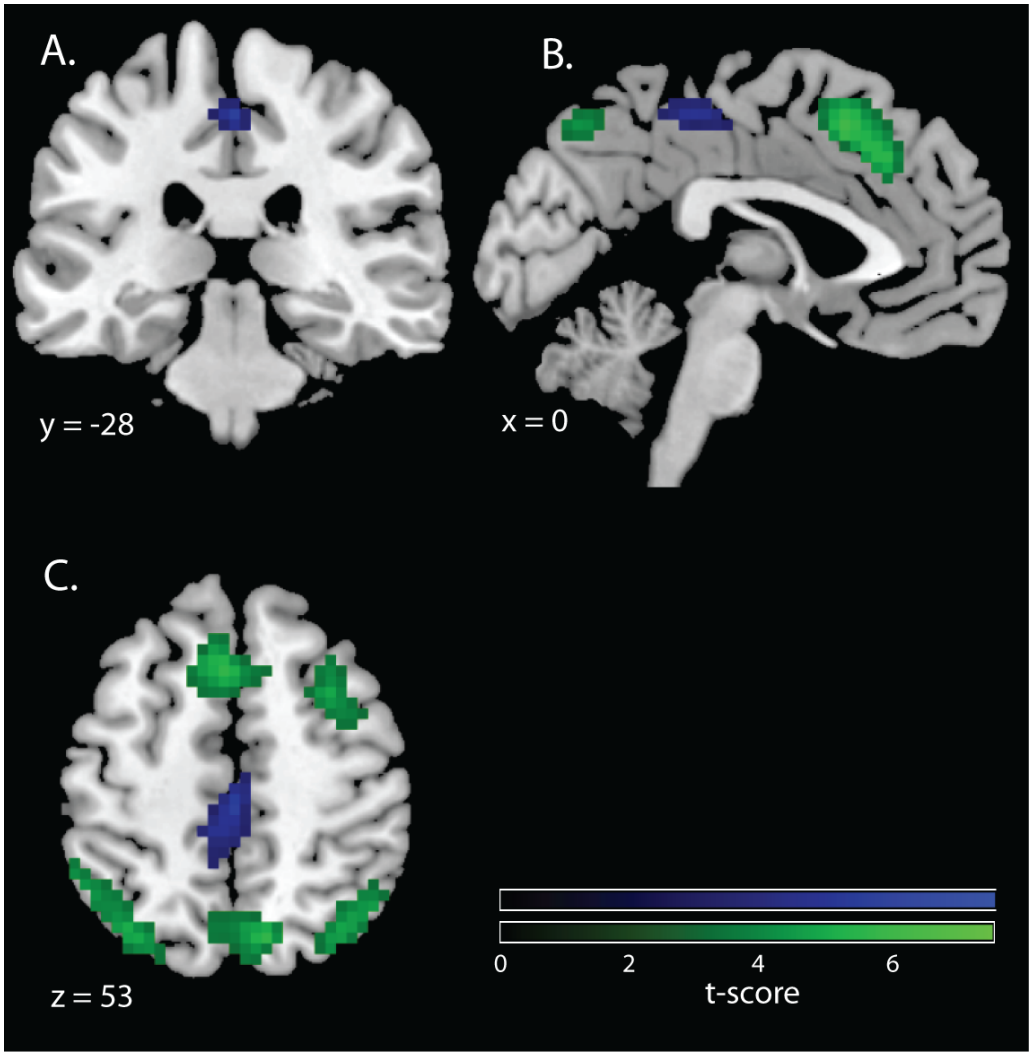


Figure 8.5: Subjective free > subjective instructed in whole brain (blue) and objective free > objective istructed (green). Panels **A,B,C** show coronal, sagittal and axial planes respectively. Blue: Postcentral region showing increased BOLD signal for the contrast free > instructed in the median split data from the subjective context. BOLD signal peaked at MNI coordinates (x = 4 y = -21 z = 49) and (x = 0 y = -28 z = 53). Results from this contrast in the subjective context are non-overlapping with those from the same contrast in the classical context (green). BOLD activations were corrected for multiple comparisons by means of a Monte Carlo simulation ($p<0.001$, minimum cluster size: 25 voxels).

8. 4 Discussion

Free action has classically been operationalized as action that is underdetermined by the external environment, and is therefore not stimulus-bound. It has been contrasted with instructed action, in which actions are fully specified by external stimuli. Although this objective operationalization does not make explicit reference to the subjective experience of acting freely, the objective and subjective freedom of action are often implicitly assumed to be related (Richard E Passingham et al., 2010; Schüür & Haggard, 2011), see for example (I. Goldberg et al., 2008).

This study investigated the relation between the classic operational definitions of free and instructed action and the subjective feeling of acting freely. Participants chose in each trial between one of four alternative actions. BOLD activations were measured in classical conditions in which these free and instructed actions were defined in the objective, classical way: either the choice of action was left to the participant (free trials), or the choice of action was fully specified in each trial (instructed trials). Results were in line with previous literature. In particular, they confirmed that medial frontal and parietal BOLD activity is associated with free actions. Results of the comparison between free and instructed conditions showed significantly greater BOLD responses for free compared with instructed in RCZ/SMA, left DLPFC, bilateral IPL and left PMC. These results are consistent with previous reports (Cunnington et al., 2006; Lau et al., 2004; Lau et al., 2006). When participants were additionally asked to rate how free their choices felt, subjective reports in this objective context were also consistent with the operational definitions of free and instructed choice, and BOLD contrasts replicated previous studies. This part of the results is broadly consistent with the classical view of voluntary action. Moreover, it confirms, seemingly for the first time, that the neural process of internal generation of action produce an experience of free selection (Krieghoff et al., 2011).

Next, this pattern of BOLD activations was compared with those obtained in a subjective context where participants selected actions according to the combination of a numerical stimulus stem and a completion rule (“look random”). The

completion rule aimed at providing each participant with a situation in which they could experience an ecologically valid graded sense of voluntariness.

It was assumed that participants might use completion rules to conform to the required "random appearance" of sequences. The precise completion rule (e.g., repetition avoidance, (Ginsburg & Karpiuk, 1994) could vary across participants according to their conceptions of randomness.

Although the completion rule was not considered to be critical in the context of this experiment, a simple analysis was carried out, to determine the factors affecting participants' number choices, and voluntariness ratings. As an initial approach, possible strategies for exclusion/inclusion strategies were sought in each participant's data. Exclusion trials were defined as those in which the number chosen was not included in the presented stem. Inclusion trials were defined as those where the number chosen was already included in the stem. Results revealed that participants tended to base their number choice on an exclusion strategy. Crucially however, the mean voluntariness ratings did not differ between inclusion and exclusion trials, suggesting that participants took into account additional factors, independent from the exclusion behavior, to evaluate the degree of voluntariness of each choice. In other words, the pattern of BOLD activity revealed by a contrast between choices subjectively rated as free vs. instructed cannot be simply explained by exclusion-related activity.

In addition, the mean voluntariness ratings were compared across all possible stimulus set sizes, where the stimulus set size was defined at the number of different numbers present in the stem. Participants felt less free in responding to sequences with low stimulus space values (e.g. "3 3 3 3"), perhaps because these sequences would have strongly precluded choosing the repeated response (e.g., "3").

In contrast, sequences with stimulus spaces of 4 (e.g., "3 2 1 4") were associated with the highest voluntariness ratings. Two possible reasons for this effect may be cited. First, such sequences may have appeared to offer no particular constraint on participants' choice. Second, stems with a stimulus space of 4 involved no number repetitions within the stem. As a result, participants could not use an exclusion rule. At the same time, responses to these stimuli were consistently rated as the most free.

Therefore, the ability to use an exclusion rule (i.e., simply choose the number that is not included in the stem) is unlikely to be the core of feeling free.

However, the actual rules used for "random" generation are not relevant for the analysis reported here, and the requirement to generate random numbers served merely to provide a plausible response space within which some responses might seem more free than others. This subjective experience was considered, independently of the precise completion rule adopted, and independently of the objective randomness of the response. Trials associated with higher subjective ratings were contrasted with those with lower subjective ratings of voluntariness.

The neural correlates of the subjective feeling of freedom were quite different from the neural correlates of free choice as classically operationalized. The ROI analysis revealed that in five of the ROIs identified (ACC, left IPL, right IPL, left DLPFC and left PMC), BOLD activity in the objective context was higher for free trials as compared with instructed trials. In stark opposition, in the subjective context BOLD activity showed was numerically lower for free trials in all of these five ROIs, and was significantly lower in two (right IPL and left PMC). Only one area, the precuneus, showed the same pattern of BOLD activity in the objective and subjective contexts. In fact, the contrast between free and instructed choices in this ROI did not reach statistical significance in either the objective or the subjective context, making this result hard to interpret. Increases in precuneus BOLD activity are not typically associated with voluntary action. Instead, the precuneus has been linked to self-referential tasks and experience of agency (Cavanna & Trimble, 2006). Intriguingly however, multivariate patterns of activity in this area can predict outcomes of free choices (Soon et al., 2008).

A whole brain fMRI analysis identified a cluster of BOLD activation in a medial postcentral area correlating with the reported subjective feelings of free>instructed choices. This did not overlap with any of the areas identified by the free > instructed contrast in the classical context. Together, these results suggest that the classic operationalization of free action and the subjective experience of choosing freely are dissociated. Therefore, classifying actions according to the subjective report of how free they feel revealed a brain area that had not been reported before in studies of volition and free action selection. The increase in BOLD activity in the medial postcentral cortex correlated with the subjective feeling of freedom in choosing

actions, but was not related to the objective informational definition of free action. Therefore, these results may argue that the feeling of acting freely in naturalistic situations may be independent of the classical objective distinction between free and instructed choice.

These results give rise to three main questions. First, why does the subjective experience of acting freely dissociate from the classical experimental manipulations of free action? Second, what is the functional relevance of the postcentral area found to be related to subjective feelings of acting freely? And finally, what is the significance of subjective report in the study of volition?

Why are the objective and subjective contexts dissociated?

To understand why the subjective experience of acting freely does not correspond to the objective operationalization of free action, it was examined which stimuli selected by participants were reported to be based on free choices, and which stimuli made them report feelings of not acting freely.

These analyses revealed that restricted "stimulus spaces" (i.e., stimulus sequences containing only a few digits, such as "1 1 1 1") were associated with low ratings of voluntariness. In contrast, sequences with large stimulus spaces were associated with high feelings of voluntariness. Presumably, participants interpreted the completion rule of keeping sequences "looking random" as forbidding simply repeating the single stimulus number. In contrast, a stimulus space with 4 items, such as the stimulus sequence "3 2 1 4" was associated with high feelings of voluntariness.

This behavioural result interestingly suggests that perceived voluntariness in our task is not related to the number of available alternatives but is instead related to how strongly the environment precludes otherwise available alternatives.

Alternative accounts

Several alternative explanations for the dissociation between objective and subjective context should be considered here. There were differences in the patterns of BOLD activity associated with the free > instructed contrast between the objective

and subjective contexts. Clearly, the two contexts differed in more than one respect. For example, the subjective context required a complex evaluation of the number sequence and may have recruited random generation processes (Jahanshahi et al., 2000). In addition, the experimental design included a memory question about the number sequences in the subjective, but the objective context did not. The subjective context involved experiences of free action choice which was clearly graded, and could vary along a continuum, while the objective context involved experiences which participants unhesitatingly recognized as located at the extreme ends of this continuum. Arguably however, the differences in BOLD activity identified between objective and subjective contexts cannot be easily explained by differences in task demands. The mental processes such as random number generation would have occurred in *both* trials felt as more free and trials felt as less free within the subjective context, but would have occurred in *none* of the trials in the objective context. Consequently, the comparison between free and instructed trials *within* each context would have subtracted out any potential BOLD activity that was directly related to differential task demands.

Alternatively, the differences in patterns of BOLD activations may be related to the nature of the free/instructed distinction in each context. There is a strong and categorical difference between free and instructed trials in the objective context. Conversely, the difference between free and instructed trials in the subjective context is slight and gradual. This alternative explanation would predict smaller differences between free and instructed conditions in the subjective context as compared to the objective context, but in the *same direction*. This was not the case. There was a trend in all ROIs identified (except the precuneus) for an *inversion* of the pattern of BOLD activation in the subjective context, as compared with the objective context. This trend became significant in two of the six ROIs analyzed, namely the right IPL and left PM. More strongly, the free > instructed contrast in the subjective context revealed significant increases in BOLD activity in the postcentral area that were not significant in the objective context.

To account for an inversion in patterns of BOLD activity, it would be necessary to appeal to an implausible inversion of the meaning of the terms “very free” and “not free” on the part of the participants. Therefore, simple differences in the nature of free and instructed contexts cannot easily account for the results reported here.

Instead, the subjective feeling of acting freely may be associated with increased BOLD activity in the postcentral area. However, other more speculative explanations cannot be ruled out. For example, activity in the medial postcentral cortex may reflect increased difficulty in finding a random-looking sequence, or other differences related to task processing such as the amount of information that was taken into consideration for the number choice.

Importantly however, previous data do not strongly support these alternative accounts. Increased task difficulty is typically associated with prefrontal, and not mediocentral areas (e.g., Barch et al., 1997).

The role of restricted choice on BOLD activity

Trials in the subjective context represent an intermediate situation between the two objective extremes of free and instructed actions. In other words, even those trials subjectively judged as instructed will require a certain degree of choice.

Two relevant studies have addressed the effect of varying degrees of choice on BOLD activity. Forstmann et al (Forstmann et al., 2006) designed a task selection experiment, in which participants were either instructed on which task to perform, or could choose the task between either two or three task alternatives. They found that BOLD activity in RCZ was higher for task choice conditions, but did not vary with the *scope* of choice (i.e., two vs. three possible tasks). This may explain why here there were no significant BOLD signal increases in the classical areas related to free choice in the subjective context. However, and importantly, limited set of response alternatives cannot readily explain the increase in BOLD activity in the medial postcentral region.

In a previous study (van Eimeren et al., 2006) participants were asked to freely select one out of a number of available response alternatives, ranging from 1 to 4. This design, which did not include a subjective judgement component, allowed the authors to compare conditions of no selection (only one available alternative), restricted selection (between 2 and 3 available alternatives) and full selection (all four available alternatives). If the medial postcentral activation is simply due to a limited availability of alternatives, then the present results should also have revealed the activations observed in the study by van Eimeren et al (van Eimeren et al., 2006)

in the contrast of restricted selection > no selection conditions. Their contrast revealed increased BOLD activity in frontal areas (rostral SMA, rostral ACC and right DLPFC), premotor areas (bilateral rostro-dorsal premotor cortex) and parietal areas (bilateral superior parietal lobule, anterior inferior parietal sulcus, bilateral parieto-occipital sulcus and bilateral cerebellum) but not in the medial postcentral cortex that are reported here. The results by Van Eimeren (2006) do not support the possibility that the BOLD activation in the medial postcentral area is simply related to the limited response space of alternatives.

What is the function of the medial postcentral region?

Could the activation in the medial postcentral area be due to a confound in the task, rather than a true correlate of experience of free choice? Here several alternative explanations are examined.

In tasks requiring free selection, human participants typically hold their previous responses in working memory. Monitoring the contents of working memory can then prevent repetition of behaviour, and can promote generation of seemingly random response sequences (Goldman & Rosvold, 1970). Consequently, working memory acts as a confound that is hard to separate from free selection. Increased working memory load has been associated with increases in BOLD activity in IPL (Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000), but not in DLPFC (Hadland et al., 2001). Given that the objective context led to relatively little cognitive load, it is possible that participants were additionally monitoring their previous history of choices when selecting a response. This would have meant increased working memory function in the objective context. On the other hand, the more cognitively demanding subjective context may have precluded working memory effects on free trials. This could explain the attenuation of BOLD activity in the ROIs analyzed. However, these results revealed not just attenuation, but an inversion of the free-instructed difference in the subjective compared to the objective context. Working memory accounts cannot easily explain this inversion.

The phenomenology of free action has rarely been addressed experimentally, despite a strong theoretical interest in the issue (Nahmias, Morris, & Nadelhoffer, 2004), but see (Haggard, Cartledge, Dafydd, & Oakley, 2004). The postcentral area that was

found to correlate with the experience of free choice has been related to resting state (Mazoyer et al., 2001) and default network functions (Mason et al., 2007). However, this area is not generally identified as a core component of the default mode network. Nevertheless, these data remain compatible with the idea that this area may be linked to reflective processes that give rise to the conscious feeling of free will (Goldberg et al., 2008).

What is the value of relying on subjective experience?

Unlike previous studies (e.g., van Eimeren et al., 2006) here it was not *assumed* that the subjective feelings of voluntariness would directly correlate with the number of available alternatives. Instead, participants were given a complex rule and were allowed to interpret it. In this way, the aim was to distil the bases of the feelings of voluntariness, *independently* from any preconceptions from the experimenter's part.

Subjective report provides unique access to phenomenal consciousness (Chalmers, 1997), but is notoriously problematic as a guide to neural bases of experience. In this vein, a recent report by Guggisberg et al (Guggisberg, Dalal, Schnider, & Nagarajan, 2011) has directly addressed the neural correlates of the subjective experience related to voluntary action. In an MEG study, the authors investigated the neural substrates of sensory events, of actual movements, and of intentions to move. In addition, they characterized the neural substrates of introspection about the timing of each of these events. They found that the spatial extent of the neural correlates of introspection of events did not match the neural loci of the first order events at which introspection was directed (but see Lau et al., 2004) for the opposite result). The present study mirrors that by Guggisberg et al (2011) in the sense that the brain areas that underlie introspection about a given process (here the feeling of free choice between action alternatives) do not appear to overlap with the brain areas that implement the the first order processes (i.e., choosing between action alternatives) to which that introspection refers.

Several other studies have directly addressed the subjective experience of intending to act. These studies have revealed an involvement of both frontal and parietal areas in the phenomenology of intended action. For example, Fried e*t al* (Fried et al., 1991) found that direct electrical stimulation to medial frontal cortex led to

subjective "urges" to act. More recently, Kühn and Haggard (Simone Kühn, Brass, & Haggard, 2012) found that BOLD activity in a caudal portion of SMA correlated with the strength of the "temporal binding" effect, which has been suggested as an implicit marker of agency (Haggard et al., 2002). On the other hand, Desmurget et al (Desmurget et al., 2009) reported the conscious experience of intending to act following direct stimulation of the parietal cortex. These studies have associated medio-frontal and parietal areas to the feeling of intended action, but not the postcentral area reported here. One important difference between those studies and the present study should be pointed out. In those studies, the timing of action was not specified, but was left up to the participant. Consequently, the subjective experience of action intention may be related to the feeling of impending action. In the present experiment, in contrast, an action was to be made at a specific time. Actions could only differ in terms of their freedom of constraint by the external environment. Therefore, the subjective experience reported here was not related to the feeling of being about to act, but rather with the feeling of internal generation of response alternatives. To use the terminology of Brass and Haggard's *what, when, whether* model (Brass & Haggard, 2008) the feeling corresponded to the "what" component of volition, not to the "when" component.

Importantly, these results do not undermine the validity of classical operational distinctions between free and instructed choice, nor do they rule out some relation between the classical operational distinction of free action and the experience of free choice. However, they do suggest that the experience of free choice of action is not simply a consequence of the same brain processes involved in free action selection. According to some theories, consciousness of a stimulus is simply a result of neural activity in particular brain areas that process information about the stimulus (Blankenburg et al., 2003; Libet, Alberts, Wright, & Feinstein, 1967). The present results suggest that the conscious experience of choosing what to do does not work in this way. Where then, does the experience of free choice arise? There are at least two possibilities. First, the feeling of freely choosing might be closer to a confabulation than to a perception. It might not be a bona fide experience at all (Wegner, 2002). Some support for this view comes from evidence on the perception of conscious intention. When people freely choose between alternative manual responses, the experience of volition appears to be generated by processes occurring

after action selection, rather than before (Haggard & Eimer, 1999). On this view, the process of selection itself might not produce any direct phenomenal consequences. Second, the feeling of voluntary choice might indeed be a genuine experience, but it would be generated by a pathway parallel to the (unconscious) pathway that generates action itself (Cleeremans, 2011; Wegner, 2002).

8. 5 Conclusion

This experiment relates to the traditional philosophical question of "free will". People generally feel that *they* choose their own actions (Nichols, 2011). Here, the aim was not to contribute to the debate over whether free will exists, or how it might be compatible with the neurobiological determinism. Rather, this study attempted to investigate the *feeling* of free will scientifically. Choosing between alternative actions in the absence of any external instruction generates a phenomenal experience that people can report (N. Block, 1996). These results suggest that this experience may not directly derive from brain circuits involved in intentional action selection, but from other brain areas. fMRI is a correlational method, so these results cannot reveal whether these other brain areas generate a subjective experience of free action choice by confabulation, or by some form of internal monitoring of action selection processes. The present results do suggest, however, that the feeling of free action choice is not straightforwardly generated by action choice itself.

Chapter 9 General discussion

This general discussion provides a brief integrative overview of the general common findings of the experiments described in this thesis. Finally, it comments on the implications of the research presented for the development of individual and social moral responsibility.

9. 1 Conclusions

Human intentional behaviour is widely recognized to have 'freedom from immediacy' (Shadlen & Gold, 2004), meaning that intentional decisions to act and inhibit action depend on the wide integration of information beyond any single current stimulus. These other factors may include, for example, memory for previous experience and predictions of future outcomes and more abstract entities such as values, moral rules, etc. However, what exactly makes intentional action intentional is still an unresolved question in the literature (See section 1.2.4).

This thesis took a critical position relative to what intentional behaviour is, and in what ways it can be distinguished from instructed behaviour. Importantly, this thesis explored how intentional behaviour is best operationally defined for its empirical study, and to what extent the classical distinction between instructed and intentional processes can be generalized from action to other aspects of behaviour, such as inhibition.

Taking this critical perspective, this approach has generated four main findings that contribute to the understanding of the mechanisms supporting intentional behaviour.

1. Intentional behaviour differs from instructed behaviour

Several previous empirical studies had compared the neural correlates of intentional and instructed actions. Chiefly on the basis of neuroimaging and lesion studies, a distinction between intentional and instructed action has been proposed, but has also been heavily criticized. Useful distinctions can be generalized beyond the domain for which they were originally developed. Therefore, and given the importance of inhibition of action in voluntary control, this thesis has attempted to generalize the instructed/intentional distinction from action to inhibition of action.

A review of the existing evidence for intentional inhibition (chapter 1) was provided, and three experimental chapters (chapters 2, 3 and 4) directly addressed potential differences between instructed and intentional inhibition. Together, the results reported here revealed that different processes occurring in the period before response execution differ between instructed and intentional behaviour, both in cases

of action and inhibition. Therefore, the results argue for a general and neurophysiologically meaningful division between two kinds of behaviour.

2. Intentional behaviour is associated with weak commitment to action decisions

Intentional decisions are allegedly the result of internal deliberation processes, and not just automatized responses to unpredictable changes in the external environment. Therefore, because intentional actions are associated with motivations for action, an intuitive account would predict that intentional behaviour will has stronger neural underpinnings than the corresponding instructed behaviour.

Results from chapters 2, 3 and 5 suggest that this intuitive account may not hold. Presumably because the alternative course of behaviour is always in principle available, intentional decisions seem to represent situations in which the counterfactual response alternative remains more active, and active until a later point in the response generation process, than instructed decisions.

A note of caution is important here. The experiments in this thesis presented participants with situations that were mostly inconsequential, and in which the outcome of their decisions would have little impact on the participants' future circumstances. It is neither practical nor efficient to study life-changing action decisions in a laboratory experiment. The recourse to study these relatively minor decisions used here may explain the "intermediate" nature of intentional behaviour. In truly naturalistic situations, the relative strengths of the neural codes associated with intentional and instructed actions may not be consistent with what has been reported here. This point is well illustrated by the results from chapter 4. Electrophysiological measures taken before the time of conscious decisions about whether to act or delay a reaction transiently revealed that unconscious preceding neural activity may impact on conscious decisions. However, and importantly, these effects could be apparently overridden by external instructions. In everyday situations, other internal or external signals could presumably override any weak effects of the previous states of the brain, that were shown to bias decisions towards one or another response alternative.

3. Representations of intentional actions are persistent

Chapter 6 of thesis revealed an interesting and unexpected corollary of the relative weakness of intentional decisions. If it is true that the representation of the chosen response alternative is weaker in cases of intentional decisions as compared to instructed decisions, this implies that the representation of the *non-selected* alternatives is also *stronger* in intentional decisions. Therefore, even after intentional response selection has taken place, the neural representation of the non-chosen alternatives is maintained, and can potentially affect behaviour. This may not be the case in situations of instructed decisions, where strong external signals provide clear and unambiguous drive for the suppression of the neural activity associated with the "incorrect" response alternatives.

4. Introspective reports may enrich models of intentional behaviour

Arguably, one of the most interesting aspects of intentional action is precisely what it *feels like* to act intentionally. In stark contrast, the traditional approach the experimental study of intentional action has favoured functional definitions of behaviour, disregarding the different experiences associated with these different behaviours. Implicit in the intentional action literature, there is the assumption that the objective definitions of intentional action faithfully capture the *subjective* experience of acting intentionally.

However this important assumption had never been challenged. The neural processes responsible for the emergence of the subjective experience of "having free will" remain largely unexplored. In chapter 8 the neural correlates of this ubiquitous subjective experience was directly addressed. Two interesting results emerged. In the first place, participants were able to consistently introspect on their feelings of voluntariness. Second, the brain correlates of intentional action identified on the basis of subjective measures did not match those that are typically identified on the basis of objective measures. These results demonstrate that subjective experience can usefully be integrated into experimental paradigms, to provide novel insights into neurocognitive questions. They also challenge the recent view that experiences, such

as urge and intention, are a direct conscious correlate of neural activity in motor areas of the brain. Instead, they point to a distinction between the information processes underlying action control and the subjective experience of action control, at least for the action selection processes studied in chapter 8.

9. 2 Social implications

As a final note, the potential social significance of understanding the neural basis of intentional behaviour should be discussed. Humans are social animals with a sophisticated pattern of interactions, based on reciprocity. All human societies have a concept of moral responsibility for action, which in addition, presupposes the capacity for intentional inhibition: the individual could have refrained from an action that they made. This socially accepted concept of intentions leads to a praise of *motu proprio*. Doing -or refraining from doing- something as a consequence of an instruction does not have the same value as doing something out of personal convictions and intentions. Thus, the understanding of what makes an action intentional, and what allows the action to be intentionally suppressed, is crucial for a full understanding of the extent of moral responsibility.

This thesis aimed to shed some light on the neurophysiological, cognitive, and subjective mechanisms involved in intentional behaviour. A better understanding of neurophysiological changes may eventually lead to changes in social concepts or morality and responsibility. However, much care should be taken to avoid direct transitions from "neural 'is' to moral 'ought'" (Greene, 2003). Other foundations are required to draw any conclusions about issues on morality and social responsibility.

One point not explored here is the development of intentional behaviour across the lifespan. Newborns act only guided instructed by the external environment and their own internal states. These internal states may be initially precarious, and related mainly to physiological needs. They become gradually more sophisticated over time. Children gradually learn to overcome prepotent behaviours e.g., (Snyder & Munakata, 2010). First, they learn to do as they are told, even if this involves going against obtaining immediate rewards. Then, children slowly start to regulate their

behaviour even in the absence of external cues, and therefore in an intentional manner.

Interestingly, a child may be irritated when she is asked or told to do something that she was planning to do in any case. This suggests that the concept of intentionality, and its distinction from instructedness, appears early in development. The question of how both the *capacity for* intentional behaviour and the *concept of* intentional behaviour may develop remains largely unexplored. Moreover, the relationship and interdependence between behavioural capacity and conceptual understanding of intention is unknown. Further, the relationship to both with the concept of self has not been explored. These questions may be difficult to address experimentally, from a neuroscientific standpoint. At the same time, they are part of what makes us human. This thesis has illustrated one way in which introspection and subjective report can be incorporated into neurophysiological paradigms, suggesting promising and innovative future research directions.

References

Aron, A. R. (2010). From Reactive to Proactive and Selective Control: Developing a Richer Model for Stopping Inappropriate Responses. *Biological Psychiatry*, *69*(12), e55–e68. doi:10.1016/j.biopsych.2010.07.024

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience*, *27*(14), 3743.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *Journal of Neuroscience*, *26*(9), 2424–2433.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *99*(20), 13172–13177. doi:10.1073/pnas.162485599

Baker, K. S., Piriyapunyaporn, T., & Cunnington, R. (2012). Neural activity in readiness for incidental and explicitly timed actions. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2011.12.026

Bancaud, J., Talairach, J., Geier, S., Bonis, A., Trottier, S., & Manrique, M. (1976). Behavioral manifestations induced by electric stimulation of the anterior cingulate gyrus in man. *Revue neurologique*, *132*(10), 705.

Banks, W. P., & Isham, E. A. (2009). We Infer Rather Than Perceive the Moment We Decided to Act. *Psychological Science*, *20*(1), 17 –21.

Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, *35*(10), 1373–1380. doi:10.1016/S0028-3932(97)00072-9

Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. *Social psychology: Handbook of basic principles*, 516–539.

Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. S. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. *The Journal of Neuroscience*, *29*(4), 1175–1190. doi:10.1523/JNEUROSCI.3328-08.2009

Beisteiner, R., Windischberger, C., Lanzenberger, R., Edward, V., Cunnington, R., Erdler, M., … Deecke, L. (2001). Finger Somatotopy in Human Motor Cortex. *NeuroImage*, *13*(6), 1016–1026. doi:10.1006/nimg.2000.0737

Bekker, E. M., Kenemans, J. L., Hoeksma, M. R., Talsma, D., & Verbaten, M. N. (2005). The pure electrophysiology of stopping. *International Journal of Psychophysiology*, *55*(2), 191–198. doi:10.1016/j.ijpsycho.2004.07.005

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*(6013), 83.

Bhushan, N., & Shadmehr, R. (1999). Computational nature of human adaptive control during learning of reaching movements in force fields. *Biological Cybernetics*, *81*(1), 39–60. doi:10.1007/s004220050543

Blankenburg, F., Taskin, B., Ruben, J., Moosmann, M., Ritter, P., Curio, G., & Villringer, A. (2003). Imperceptible Stimuli and Sensory Processing Impediment. *Science*, *299*(5614), 1864–1864. doi:10.1126/science.1080806

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., & Linden, D. E. J. (2004). Localizing P300 Generators in Visual Target and Distractor Processing: A Combined Event-Related Potential and Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, *24*(42), 9353–9360. doi:10.1523/JNEUROSCI.1897-04.2004

Block, N. (1996). On A Confusion About a Function of Consciousness.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*(3), 330–343. doi:10.1016/j.actpsy.2010.03.006

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron*, *62*(5), 733–743. doi:10.1016/j.neuron.2009.05.014

Boulinguez, P., Ballanger, B., Granjon, L., & Benraiss, A. (2009). The paradoxical effect of warning on reaction time: Demonstrating proactive response inhibition with event-related potentials. *Clin Neurophysiol*, *120*(4), 730–737.

Boulinguez, P., Jaffard, M., Granjon, L., & Benraiss, A. (2008). Warning signals induce automatic EMG activations and proactive volitional inhibition: evidence from analysis of error distribution in simple RT. *Journal of Neurophysiology*, *99*(3), 1572–1578. doi:10.1152/jn.01198.2007

Bradshaw, J. L. (2001). Asymmetries in preparation for action. *Trends in Cognitive Sciences*, *5*(5), 184–185.

Bramhall, J. (1655). A Defence of True Liberty from Ante-cedent and Extrinsecall Necessity, being an answer to a late book of Mr. *Thomas Hobbes of Malmsbury, intituled, A Treatise of Liberty and Necessity*.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *Journal of Neuroscience*, *27*(34), 9141.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *14*(4), 319–325.

Britz, J., & Michel, C. M. (2010). Errors can be related to pre-stimulus differences in ERP topography and their concomitant sources. *NeuroImage*, *49*(3), 2774–2782. doi:10.1016/j.neuroimage.2009.10.033

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The Phase of Ongoing EEG Oscillations Predicts Visual Perception. *The Journal of Neuroscience*, *29*(24), 7869–7876. doi:10.1523/JNEUROSCI.0113-09.2009

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048–16053. doi:10.1073/pnas.1004801107

Cauda, F., Geminiani, G., D'Agata, F., Duca, S., & Katiuscia, S. (2011). Discovering the somatotopic organization of the motor areas of the medial wall using low-frequency bold fluctuations. *Human Brain Mapping*, *32*(10), 1566–1579. doi:10.1002/hbm.21132

Cavanna, A. E., & Trimble, M. R. (2006). The Precuneus: A Review of Its Functional Anatomy and Behavioural Correlates. *Brain*, *129*(3), 564–583.

Chalmers, D. J. (1997). Facing up to the problem of consciousness. *Explaining Consciousness: The Hard Problem*, 9–32.

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience & Biobehavioral Reviews*, *33*(5), 631–646. doi:10.1016/j.neubiorev.2008.08.016

Chauvel, P. Y., Rey, M., Buser, P., & Bancaud, J. (1996). What stimulation of the supplementary motor area in humans tells about its functional organization. *Advances in neurology*, *70*, 199–209.

Cheyne, J., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98–113. doi:10.1016/j.cognition.2008.12.009

Chikazoe, J. (2010). Localizing performance of go/no-go tasks to prefrontal cortical subregions. *Current Opinion in Psychiatry*, *23*(3), 267.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585–1599. doi:10.1098/rstb.2007.2054

Cisek, P., & Kalaska, J. F. (2005). Neural Correlates of Reaching Decisions in Dorsal Premotor Cortex: Specification of Multiple Direction Choices and Final Selection of Action. *Neuron*, *45*(5), 801–814. doi:10.1016/j.neuron.2005.01.027

Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to be Conscious. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00086

Cothran, D. L., & Larsen, R. (2008). Comparison of inhibition in two timed reaction tasks: the color and emotion Stroop tasks. *The Journal of Psychology: Interdisciplinary and Applied*, *142*(4), 373–385.

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, *24*(4), 349.

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. *Neuroimage*, *15*(2), 373–385.

Cunnington, R., Windischberger, C., & Moser, E. (2005). Premovement activity of the pre-supplementary motor area and the readiness for action: Studies of time-resolved event-related functional MRI. *Human Movement Science*, *24*(5-6), 644–656. doi:10.1016/j.humov.2005.10.001

Cunnington, R., Windischberger, C., Robinson, S., & Moser, E. (2006). The selection of intended actions and the observation of others' actions: a time-resolved fMRI study. *NeuroImage*, *29*(4), 1294–1302.

Daniels, C., Witt, K., Wolff, S., Jansen, O., & Deuschl, G. (2003). Rate dependency of the human cortical network subserving executive functions during generation of random number series – a functional magnetic resonance imaging study. *Neuroscience Letters*, *345*(1), 25–28.

De Jong, R., Coles, M. G., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 164–182. doi:10.1037/0096-1523.16.1.164

De Jong, R., Gladwin, T. E., & Hart, B. M. (2006). Movement-related EEG indices of preparation in task switching and motor control. *Brain research*, *1105*(1), 73–82. doi:10.1016/j.brainres.2006.03.030

Debaere, F., Wenderoth, N., Sunaert, S., Van Hecke, P., & Swinnen, S. P. (2003). Internal vs external generation of movements: differential neural pathways involved in bimanual coordination performed in the presence or absence of augmented visual feedback. *Neuroimage*, *19*(3), 764–776.

Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schürhoff, F., … Martinot, J. L. (2003). Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23), 13722.

Deiber, M. P., Passingham, R. E., Colebatch, J., Friston, K., Nixon, P., & Frackowiak, R. (1991). Cortical areas and the selection of movement: a study with positron emission tomography. *Experimental Brain Research*, *84*(2), 393–402.

Della Sala, S., & Marchetti, C. (2005). Anarchic hand. *Higher-order Motor Disorders: From Neuroanatomy and Neurobiology to Clinical Neurology*, 291.

Demiralp, T., Ademoglu, A., Comerchero, M., & Polich, J. (2001). Wavelet Analysis of P3a and P3b. *Brain Topography*, *13*(4), 251–267. doi:10.1023/A:1011102628306

Demiralp, T., Ademoglu, A., Schürmann, M., Basar-Eroglu, C., & Basar, E. (1999). Detection of P300 Waves in Single Trials by the Wavelet Transform (WT). *Brain and Language*, *66*(1), 108–128. doi:10.1006/brln.1998.2027

Desmet, C., Fias, W., Hartstra, E., & Brass, M. (2011). Errors and Conflict at the Task Level and the Response Level. *The Journal of Neuroscience*, *31*(4), 1366–1374.

Desmond, J. E., Gabrieli, J. D. E., & Glover, G. H. (1998). Dissociation of Frontal and Cerebellar Activity in a Cognitive Task: Evidence for a Distinction between Selection and Search* 1. *Neuroimage*, *7*(4), 368–376.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement Intention After Parietal Cortex Stimulation in Humans. *Science*, *324*(5928), 811.

Dettmers, C., Fink, G. R., Lemon, R. N., Stephan, K. M., Passingham, R. E., Silbersweig, D., … Frackowiak, R. S. (1995). Relation Between Cerebral Activity and Force in the Motor Areas of the Human Brain. *Journal of Neurophysiology*, *74*(2), 802–815.

Dimoska, A., Johnstone, S. J., & Barry, R. J. (2006). The auditory-evoked N2 and P3 components in the stop-signal task: Indices of inhibition, response-conflict or error-detection? *Brain and Cognition*, *62*(2), 98–112. doi:10.1016/j.bandc.2006.03.011

Dirnberger, G., Duregger, C., Lindinger, G., & Lang, W. (2011). On the regularity of preparatory activity preceding movements with the dominant and non-dominant hand: a readiness potential study. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *81*(2), 127–131. doi:10.1016/j.ijpsycho.2011.04.008

Dirnberger, G., Fickel, U., Lindinger, G., Lang, W., & Jahanshahi, M. (1998). The mode of movement selection Movement-related cortical potentials prior to freely selected and repetitive movements. *Experimental Brain Research*, *120*(2), 263–272.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? Retrieved from http://psycnet.apa.org/psycinfo/1989-21502-001

Donkers, F. C. L., & Van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition*, *56*(2), 165–176. doi:10.1016/j.bandc.2004.04.005

Duann, J.-R., Ide, J. S., Luo, X., & Li, C. R. (2009). Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(32), 10171–10179. doi:10.1523/JNEUROSCI.1300-09.2009

Dum, R., & Strick, P. (1991). The origin of corticospinal projections from the premotor areas in the frontal lobe. *The Journal of neuroscience*, *11*(3), 667.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*(1–2), 1–52. doi:10.1016/0301-0511(82)90016-3

Duvernoy, H. M. (1991). The human brain: Structure, three-dimensional sectional anatomy and MRI. *Wien: Springer-Verlag*.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136.

Ebert, J. P., & Wegner, D. (2011). Mistaking randomness for free will. *Consciousness and Cognition*, *20*, 965–971.

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biological Psychology*, *35*(2), 123–138.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240. doi:10.1037/0096-1523.27.1.229

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica*, *101*(2-3), 267–291. doi:16/S0001-6918(99)00008-6

Filevich, E., & Haggard, P. (2012). Grin and bear it! Neural consequences of a voluntary decision to act or inhibit action. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *223*(3), 341–351. doi:10.1007/s00221-012-3263-9

Filevich, E., Kühn, s, & Haggard, P. (2012a). Intentional inhibition in human action: The power of "no". *Neuroscience and Biobehavioral Reviews*, *36*(4), 1107–1118.

Filevich, E., Kühn, S., & Haggard, P. (2012b). Negative motor Phenomena in cortical stimulation: implications for inhibitory control of human action. *Cortex*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010945212001414

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5*(7), 1688–1703.

Fleming, S. M., Mars, R. B., Gladwin, T. E., & Haggard, P. (2009). When the Brain Changes Its Mind: Flexibility of Action Selection in Instructed and Free Choices. *Cerebral Cortex*, *19*(10), 2352 –2360. doi:10.1093/cercor/bhn252

Fletcher, P. C., Shallice, T., & Dolan, R. J. (2000). "Sculpting the Response Space"—An Account of Left Prefrontal Activation at Encoding. *NeuroImage*, *12*(4), 404–417. doi:10.1006/nimg.2000.0633

Forstmann, B. U., Brass, M., Koch, I., & Von Cramon, D. Y. (2006). Voluntary Selection of Task Sets Revealed by Functional Magnetic Resonance Imaging. *Journal of Cognitive Neuroscience*, *18*(3), 388–398. doi:10.1162/jocn.2006.18.3.388

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007). Intrinsic Fluctuations within Cortical Systems Account for Intertrial Variability in Human Behavior. *Neuron*, *56*(1), 171–184. doi:10.1016/j.neuron.2007.08.023

Fried, I., Katz, A., McCarthy, G., Sass, K., Williamson, P., Spencer, S., & Spencer, D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *J. Neurosci.*, *11*(11), 3656–3666.

Frith, C. D., Friston, K., Liddle, P. F., & Frackowiak, R. S. (1991). Willed action and the prefrontal cortex in man: a study with PET. *Proceedings. Biological Sciences / The Royal Society*, *244*(1311), 241–246.

Frith, C., Gallagher, H., & Maguire, E. (2004). Mechanisms of control. Retrieved from http://eprints.ucl.ac.uk/128459/

Fulton, J., & Jacobsen, C. (1935). The functions of the frontal lobes: A comparative study in monkeys, chimpanzees and man. *Abstracts of the Second International Neurological Congress*, *4*, 70–71.

Ginsburg, N., & Karpiuk, P. (1994). Random generation: Analysis of the responses. *Perceptual and Motor Skills*, *79*(3 Pt 1), 1059–1067.

Gladwin, T. E., Lindsen, J. P., & De Jong, R. (2006). Pre-stimulus EEG effects related to response speed, task switching and upcoming response hand. *Biological psychology*, *72*(1), 15–34.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, *5*(1), 10–16.

Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behavioral and Brain Sciences*, *8*(04), 567–588.

Goldberg, I., Ullman, S., & Malach, R. (2008). Neuronal correlates of "free will" are associated with regional specialization in the human intrinsic/default network. *Consciousness and Cognition*, *17*(3), 587–601.

Goldman, P. S., & Rosvold, H. E. (1970). Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Experimental Neurology*, *27*(2), 291–304.

Goldman-Rakic, P. s. (2007). Motor Control Function of the Prefrontal Cortex. In G. B. Organizer, eve O'Connor, & Joanrsh (Eds.), *Ciba Foundation Symposium 132 - Motor Areas of the Cerebral Cortex* (pp. 187–200). John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470513545.ch12/summary

Goldstein, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and novelty P3. *Psychophysiology*, *39*(6), 781–790.

Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition*, *11*(2), 191–197.

Gratton, G., Bosco, C. M., Kramer, A. F., Coles, M. G. H., Wickens, C. D., & Donchin, E. (1990). Event-related brain potentials as indices of information extraction and response priming. *Electroencephalography and Clinical Neurophysiology*, *75*(5), 419–432. doi:10.1016/0013-4694(90)90087-Z

Greene, J. (2003). From neural "is" to moral "ought": what are the moral implications of neuroscientific moral psychology? *Nature*, *4*, 847–850.

Gruber, M. J., & Otten, L. J. (2010). Voluntary control over prestimulus activity related to encoding. *The Journal of Neuroscience*, *30*(29), 9793.

Guggisberg, A. G., Dalal, S. S., Schnider, A., & Nagarajan, S. S. (2011). The neural basis of event-time introspection. *Consciousness and Cognition*, *20*(4), 1899–1915.

Gumenyuk, V., Korzyukov, O., Alho, K., Escera, C., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Brain activity index of distractibility in normal school-age children. *Neuroscience Letters*, *314*(3), 147–150. doi:10.1016/S0304-3940(01)02308-4

Guye, M., Parker, G. J. ., Symms, M., Boulby, P., Wheeler-Kingshott, C. A. ., Salek-Haddadi, A., … Duncan, J. S. (2003). Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *NeuroImage*, *19*(4), 1349–1360. doi:10.1016/S1053-8119(03)00165-4

Habas, C. (2010). Functional connectivity of the human rostral and caudal cingulate motor areas in the brain resting state at 3T. *Neuroradiology*, *52*(1), 47–59. doi:10.1007/s00234-009-0572-1

Hadland, K. A., Rushworth, M. F. S., Passingham, R. E., Jahanshahi, M., & Rothwell, J. C. (2001). Interference with Performance of a Response Selection Task that has no Working Memory Component: An rTMS Comparison of the Dorsolateral Prefrontal and Medial Frontal Cortex. *Journal of Cognitive Neuroscience*, *13*(8), 1097–1108.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934–946.

Haggard, P., Cartledge, P., Dafydd, M., & Oakley, D. A. (2004). Anomalous control: when "free-will"is not conscious. *Consciousness and Cognition*, *13*(3), 646–654.

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

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *126*(1), 128–133.

Hagura, N., Kanai, R., Orgs, G., & Haggard, P. (2012). Ready steady slow: action preparation slows the subjective passage of time. *Proceedings of the Royal Society B: Biological Sciences*. doi:10.1098/rspb.2012.1339

Handwerker, H. O. (2010). Microneurography of pruritus. *Neuroscience Letters*, *470*(3), 193–196. doi:10.1016/j.neulet.2009.06.092

He, B., Lian, J., Spencer, K. M., Dien, J., & Donchin, E. (2001). A cortical potential imaging analysis of the P300 and novelty P3 components. *Human brain mapping*, *12*(2), 120–130.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Hikosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, *12*(2), 217–222. doi:10.1016/S0959-4388(02)00307-0

Hillyard, S. A. (1969). Relationships between the contingent negative variation (CNV) and reaction time. *Physiology & Behavior*, *4*(3), 351–357.

Hluštík, P., Solodkin, A., Gullapalli, R. P., Noll, D. C., & Small, S. L. (2001). Somatotopy in Human Primary Motor and Somatosensory Hand Representations Revisited. *Cerebral Cortex*, *11*(4), 312–321. doi:10.1093/cercor/11.4.312

Hobbes, T. (1937). *Leviathan*. Retrieved from http://books.google.com/books?hl=en&lr=&id=7M1ltciLaU8C&oi=fnd&pg=PT762&dq=hobbes+leviathan&ots=r8Qzo4MFCw&sig=9QXMLSc8GXoVkkQa0WCrH8lq2Ns

Hobbes, T., & Bramhall, J. (1999). *Hobbes and Bramhall on Liberty and Necessity*. Cambridge University Press.

Hoffstaedter, F., Grefkes, C., Zilles, K., & Eickhoff, S. B. (2012). The "What" and "When" of Self-Initiated Movements. *Cerebral Cortex*. doi:10.1093/cercor/bhr391

Hoshi, E., & Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Current Opinion in Neurobiology*, *17*(2), 234–242. doi:10.1016/j.conb.2007.02.003

Hughes, G., Schütz-Bosbach, S., & Waszak, F. (2011). One Action System or Two? Evidence for Common Central Preparatory Mechanisms in Voluntary and Stimulus-Driven Actions. *The Journal of Neuroscience*, *31*(46), 16692 –16699. doi:10.1523/JNEUROSCI.2256-11.2011

Hughes, G., Velmans, M., & Fockert, J. D. (2009). Unconscious priming of a no-go response. *Psychophysiology*, *46*(6), 1258–69.

Ikeda, A., Hirasawa, K., Kinoshita, M., Hitomi, T., Matsumoto, R., Mitsueda, T., … Takahashi, R. (2009). Negative motor seizure arising from the negative motor

area: Is it ictal apraxia? *Epilepsia*, *50*(9), 2072–2084. doi:10.1111/j.1528-1167.2009.02097.x

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*(7), 273–280. doi:10.1016/j.tics.2008.04.002

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*(2), 139.

Jahanshahi, M., Dirnberger, G., Fuller, R., & Frith, C. (2000). The role of the dorsolateral prefrontal cortex in random number generation: a study with positron emission tomography. *Neuroimage*, *12*(6), 713–725.

Jahanshahi, M., Jenkins, I. H., Brown, R. G., Marsden, C. D., Passingham, R. E., & Brooks, D. J. (1995). Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain*, *118*(4), 913–933.

James, W. (1890). The principles of psychology, Vol I.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature neuroscience*, *8*(2), 234–241.

Jenkins, I. H., Jahanshahi, M., Jueptner, M., Passingham, R. E., & Brooks, D. J. (2000). Self-initiated versus externally triggered movements. *Brain*, *123*(6), 1216 –1228. doi:10.1093/brain/123.6.1216

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, *82*(6), 477–482. doi:10.1016/0013-4694(92)90054-L

Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., … Matthews, P. M. (2004). Changes in Connectivity Profiles Define Functionally Distinct Regions in Human Medial Frontal Cortex. *Proceedings of*

the National Academy of Sciences of the United States of America, *101*(36), 13335–13340. doi:10.1073/pnas.0403743101

Joordens, S., Van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: biases in the measurement of voluntary actions. *Consciousness and Cognition*, *11*(2), 231–240; discussion 308–313.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. s. j., & Passingham, R. E. (1997). Anatomy of Motor Learning. II. Subcortical Structures and Learning by Trial and Error. *Journal of Neurophysiology*, *77*(3), 1325–1337.

Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and Movement Representations in the Primary Motor Cortex. *Science*, *285*(5436), 2136–2139. doi:10.1126/science.285.5436.2136

Kane, R. (2005). *A Contemporary Introduction to Free Will*. Oxford University Press, USA.

Karch, S., Mulert, C., Thalmeier, T., Lutz, J., Leicht, G., Meindl, T., … Pogarell, O. (2009). The free choice whether or not to respond after stimulus presentation. *Human Brain Mapping*, *30*(9), 2971–2985. doi:10.1002/hbm.20722

Kennerley, S. W., Sakai, K., & Rushworth, M. f. s. (2004). Organization of Action Sequences and the Role of the Pre-SMA. *Journal of Neurophysiology*, *91*(2), 978–993. doi:10.1152/jn.00651.2003

Kitamura, J., Shibasaki, H., & Kondo, T. (1993). A cortical slow potential is larger before an isolated movement of a single finger than simultaneous movement of two fingers. *Electroencephalography and clinical Neurophysiology*, *86*(4), 252–258.

Kitamura, J., Shibasaki, H., Takagi, A., Nabeshima, H., & Yamaguchi, A. (1993). Enhanced negative slope of cortical potentials before sequential as compared with simultaneous extensions of two fingers. *Electroencephalography and clinical neurophysiology*, *86*(3), 176–182.

Klein-Flügge, M. C., & Bestmann, S. (2012). Time-Dependent Changes in Human Corticospinal Excitability Reveal Value-Based Competition for Action during Decision Processing. *The Journal of Neuroscience*, *32*(24), 8373–8382. doi:10.1523/JNEUROSCI.0270-12.2012

Koechlin, E., Corrado, G., Pietrini, P., & Grafman, J. (2000). Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *Proceedings of the National Academy of Sciences*, *97*(13), 7651 –7656. doi:10.1073/pnas.130177397

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. doi:10.1126/science.1088545

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235. doi:10.1016/j.tics.2007.04.005

Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P. ., & Ridderinkhof, K. R. (2004). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology*, *41*(1), 9–20.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotential\änderungen bei Willk\ürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pfl\ügers Archiv European Journal of Physiology*, *284*(1), 1–17.

Krieghoff, V., Brass, M., Prinz, W., & Waszak, F. (2009). Dissociating what and when of intentional actions. *Frontiers in Human Neuroscience*, *3*, 3.

Krieghoff, V., Waszak, F., Prinz, W., & Brass, M. (2011). Neural and behavioral correlates of intentional actions. *Neuropsychologia*, *49*(5), 767–776.

Kühn, S, & Brass, M. (2010). The cognitive representation of intending not to act: Evidence for specific non-action-effect binding. *Cognition*, *117*(1), 9–16. doi:10.1016/j.cognition.2010.06.006

Kühn, S, Elsner, B., Prinz, W., & Brass, M. (2009). Busy doing nothing: Evidence for nonaction—effect binding. *Psychonomic Bulletin & Review*, *16*(3), 542.

Kühn, S, Haggard, P., & Brass, M. (2009). Intentional inhibition: How the "veto-area" exerts control. *Human Brain Mapping*, *30*(9), 2834–2843. doi:10.1002/hbm.20711

Kühn, S., & Brass, M. (2009). When doing nothing is an option: the neural correlates of deciding whether to act or not. *Neuroimage*, *46*(4), 1187–1193.

Kühn, S., Gevers, W., & Brass, M. (2009). The neural correlates of intending not to do something. *Journal of Neurophysiology*, *101*(4), 1913.

Kühn, Simone, Bodammer, N. C., & Brass, M. (2010). Dissociating mental states related to doing nothing by means of fMRI pattern classification. *NeuroImage*, *53*(4), 1294–1300. doi:10.1016/j.neuroimage.2010.07.006

Kühn, Simone, Brass, M., & Haggard, P. (2012). Feeling in control: Neural correlates of experience of agency. *Cortex; a journal devoted to the study of the nervous system and behavior*. doi:10.1016/j.cortex.2012.09.002

Lang, W. (2003). Surface recordings of the Bereitschaftspotential in normals. *The Bereitschaftspotential: Movement-Related Cortical Potentials. New York, NY: Kluwer Academics/Plenum Publishers*, 19–34.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). *Attention to intention* (Vol. 303). American Association for the Advancement of Science.

Lau, H., Rogers, R. ., Ramnani, N., & Passingham, R. . (2004). Willed action and attention to the selection of action. *NeuroImage*, *21*(4), 1407–1415. doi:10.1016/j.neuroimage.2003.10.034

Lau, H., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science*, *303*(5661), 1208–1210.

Lau, H., Rogers, R. D., & Passingham, R. E. (2006a). Dissociating response selection and conflict in the medial frontal surface. *Neuroimage*, *29*(2), 446–451.

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

Leon, M. I., & Shadlen, M. N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*, *38*(2), 317–327.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences*, *1224*(1), 40–62. doi:10.1111/j.1749-6632.2011.05958.x

Lewis, P. A., & Miall, R. C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia*, *41*(12), 1583–1592.

Libet, B. (1985). Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *Behavioral and Brain Sciences*, *8*(04), 529–539.

Libet, B. (1999). Do we have free will. *Journal of Consciousness Studies*, *6*, 47–57.

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

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*(3808), 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 activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain: A Journal of Neurology*, *106 (Pt 3)*, 623–642.

Libet, B., Wright Jr., E. ., & Gleason, C. . (1982). Readiness-potentials preceding unrestricted “spontaneous” vs. pre-planned voluntary acts. *Electroencephalography and Clinical Neurophysiology*, *54*(3), 322–335.

Logan, G. D. (1994). On the ability to inhibit thought and action: A users’ guide to the stop signal paradigm. *Inhibitory processes in attention, memory and language*, 189–239.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: a model and a method. *Journal of Experimental Psychology*, *10*(2), 276–291.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Lüders, H. O., Dinner, D. S., Morris, H. H., Wyllie, E., & Comair, Y. G. (1995). Cortical electrical stimulation in humans. The negative motor areas. *Advances in neurology*, *67*, 115.

Masaki, H., Takasawa, N., & Yamazaki, K. (1998). Enhanced negative slope of the readiness potential preceding a target force production task. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(4), 390–397.

Mason, M. F., Norton, M. I., Horn, J. D. V., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science*, *315*(5810), 393–395. doi:10.1126/science.1131295

Matsuhashi, M., & Hallett, M. (2008). The timing of the conscious intention to move. *European Journal of Neuroscience*, *28*(11), 2344.

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as

defined by functional brain imaging: a meta-analysis. *Neuroimage*, *31*(4), 1453–1474.

Mayr, U. (2004). Conflict, consciousness, and control. *Trends in Cognitive Sciences*, *8*(4), 145–148. doi:10.1016/j.tics.2004.02.006

Mazaheri, A., DiQuattro, N. E., Bengson, J., & Geng, J. J. (2011). Pre-Stimulus Activity Predicts the Winner of Top-Down vs. Bottom-Up Attentional Selection. *PLoS ONE*, *6*(2), e16243. doi:10.1371/journal.pone.0016243

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., … Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin*, *54*(3), 287–298. doi:10.1016/S0361-9230(00)00437-8

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological bulletin*, *126*(6), 910.

Michotte, A., Thines, G., & Crabbe, G. (1991). Amodal completion of perceptual structures. *Michotte's experimental phenomenology of perception*, 140–167.

Mikuni, N., Ohara, S., Ikeda, A., Hayashi, N., Nishida, N., Taki, J., … Hashimoto, N. (2006). Evidence for a wide distribution of negative motor areas in the perirolandic cortex. *Clinical Neurophysiology*, *117*(1), 33–40.

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202. doi:10.1146/annurev.neuro.24.1.167

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature neuroscience*, *8*(7), 950–954.

Mostofsky, S. H., & Simmonds, D. J. (2008). Response Inhibition and Response Selection: Two Sides of the Same Coin. *Journal of Cognitive Neuroscience*, *20*(5), 751–761. doi:10.1162/jocn.2008.20500

Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, *37*(4), 1354–1361.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nat Rev Neurosci*, *5*(3), 218–228. doi:10.1038/nrn1345

Nachev, P. (2010). Free choice and voluntary action. *Phsyche*, *16*(1), 87–98.

Nachev, P., & Husain, M. (2010). Action and the fallacy of the'internal': Comment on Passingham et al. *Trends in cognitive sciences*, *14*(5), 192.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat Rev Neurosci*, *advanced online publication*. doi:10.1038/nrn2478

Nachev, P., Rees, G., Parton, A., Kennard, C., & Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Current Biology*, *15*(2), 122–128.

Nagai, Y., Critchley, H. D., Featherstone, E., Fenwick, P. B. C., Trimble, M. R., & Dolan, R. J. (2004). Brain activity relating to the contingent negative variation: an fMRI investigation. *Neuroimage*, *21*(4), 1232–1241.

Nahmias, E., Morris, S., & Nadelhoffer, T. (2004). The Phenomenology of Free Will. *Journal of Consciousness Studies*, *11*(7-8), 162–179.

Nathaniel-James, D., & Frith, C. (2002). The role of the dorsolateral prefrontal cortex: evidence from the effects of contextual constraint in a sentence completion task. *Neuroimage*, *16*(4), 1094–1102.

Nichols, S. (2011). Experimental Philosophy and the Problem of Free Will. *Science*, *331*(6023), 1401.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*(2), 330–357.

Nickerson, R. S., & Butler, S. F. (2009). On producing random binary sequences. *The American Journal of Psychology*, *122*(2), 141–151.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. doi:10.1038/nn.2886

Obhi, S. S., & Haggard, P. (2004). Internally generated and externally triggered actions are physically distinct and independently controlled. *Experimental Brain Research*, *156*(4), 518–523.

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature neuroscience*, *9*(4), 489–491.

Pariyadath, V., & Eagleman, D. (2007). The Effect of Predictability on Subjective Duration. *PLoS ONE*, *2*(11), e1264. doi:10.1371/journal.pone.0001264

Park, J., Schlag-Rey, M., & Schlag, J. (2003). Voluntary action expands perceived duration of its sensory consequence. *Experimental Brain Research*, *149*(4), 527–529. doi:10.1007/s00221-003-1376-x

Parkinson, J., Springer, A., & Prinz, W. (2011). Can you see me in the snow? Action simulation aids the detection of visually degraded human motion. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1463–1472. doi:10.1080/17470218.2011.594895

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, *33*(3), 1065–1076.

Passingham, R E. (1985). Premotor cortex: sensory cues and movement. *Behavioural Brain Research*, *18*(2), 175–185.

Passingham, R. E. (1987). Two cortical systems for directing movement. *Motor areas of the cerebral cortex*, *132*. Retrieved from

http://books.google.com/books?hl=en&lr=&id=LZjQjtUfo7UC&oi=fnd&pg=PA151&dq=passingham+ciba+foundation&ots=cedFALqapL&sig=lyFjP2R6uvPb3cPTndtc08FIqrE

Passingham, R. E., Bengtsson, S. L., & Lau, H. (2009). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in cognitive sciences*.

Passingham, R. E., Bengtsson, S. L., & Lau, H. (2010). Is it fallacious to talk of self-generated action?: Response to Nachev and Husain. *Trends in Cognitive Sciences*, *14*(5), 193–194. doi:10.1016/j.tics.2010.03.004

Passingham, Richard E, Bengtsson, S. L., & Lau, H. (2010). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in cognitive sciences*, *14*(1), 16–21. doi:10.1016/j.tics.2009.11.001

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389–443.

Penfield, W., & Welch, K. (1951). The supplementary motor area of the cerebral cortex: a clinical and experimental study. Arch. Neural. *Psychiatry*, *66*, 289–317.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, *60*(5), 423–434. doi:16/0013-4694(85)91017-X

Picard, N., & Strick, P. L. (1996). Motor Areas of the Medial Wall: A Review of Their Location and Functional Activation. *Cerebral Cortex*, *6*(3), 342–353. doi:10.1093/cercor/6.3.342

Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology*, *11*(6), 663–672. doi:10.1016/S0959-4388(01)00266-5

Polich, J., & Ochoa, C. J. (2004). Alcoholism risk, tobacco smoking, and P300 event-related potential. *Clinical Neurophysiology*, *115*(6), 1374–1383. doi:10.1016/j.clinph.2004.01.026

Posner, M. I. (2005). Timing the Brain: Mental Chronometry as a Tool in Neuroscience. *PLoS Biol*, *3*(2), e51.

Praamstra, P., Stegeman, D. F., Horstink, M., Brunia, C. H. M., & Cools, A. R. (1995). Movement-related potentials preceding voluntary movement are modulated by the mode of movement selection. *Experimental brain research*, *103*(3), 429–439.

Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From Salience to Saccades: Multiple-Alternative Gated Stochastic Accumulator Model of Visual Search. *The Journal of Neuroscience*, *32*(10), 3433–3446. doi:10.1523/JNEUROSCI.4622-11.2012

Rakos, R. F., Laurene, K. R., Skala, S., & Slane, S. (2008). Belief in free will: Measurement and conceptualization innovations. *Behavior and Social Issues*, *17*(1), 20–39.

Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: The N2/P3 complex further validated. *Brain and Cognition*, *56*(2), 234–252. doi:10.1016/j.bandc.2004.07.002

Richard Clark, C., McFarlane, A. C., Weber, D. L., & Battersby, M. (1996). Enlarged frontal P300 to stimulus change in panic disorder. *Biological Psychiatry*, *39*(10), 845–856. doi:10.1016/0006-3223(95)00288-X

Rigoni, D., Kühn, S., Sartori, G., & Brass, M. (2011). Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation: The Brain Minds Whether We Believe in Free Will or Not. *Psychological Science: A Journal of the American Psychological Society / APS*.

Ringkamp, M., Schepers, R. J., Shimada, S. G., Johanek, L. M., Hartke, T. V., Borzan, J., … Meyer, R. A. (2011). A Role for Nociceptive, Myelinated Nerve Fibers in Itch

Sensation. *The Journal of Neuroscience*, *31*(42), 14841–14849. doi:10.1523/JNEUROSCI.3005-11.2011

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and "top-top"control of action in cognitive experiments. *Psychological Research*, *68*(2), 189–198.

Rosenbaum, M. (1980). A schedule for assessing self-control behaviors: Preliminary findings. *Behavior Therapy*, *11*(1), 109–121.

Rosenberg-Katz, K., Singer, N., Podlipsky, I., Intrator, N., Fried, I., & Hendler, T. (2012). Enhanced functional synchronization of medial and lateral PFC underlies internally-guided action planning. *Frontiers in Human Neuroscience*, *6*, 79. doi:10.3389/fnhum.2012.00079

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, *27*(3), 832–837.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust1. *Journal of personality*, *35*(4), 651–665.

Roux, S., Coulmance, M., & Riehle, A. (2003). Context-related representation of timing processes in monkey motor cortex. *European Journal of Neuroscience*, *18*(4), 1011–1016. doi:10.1046/j.1460-9568.2003.02792.x

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The Prefrontal Cortex: Response Selection or Maintenance Within Working Memory? *Science*, *288*(5471), 1656 –1660.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., … Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations

across different versions of go/no-go and stop tasks. *NeuroImage*, *13*(2), 250–261. doi:10.1006/nimg.2000.0685

S. A. S. Institute. (1999). *SAS/STAT User's guide: Version 8* (Vol. 1). SAS institute.

Sakai, K., Hikosaka, O., Miyauchi, S., Sasaki, Y., Fujimaki, N., & Pütz, B. (1999). Presupplementary motor area activation during sequence learning reflects visuo-motor association. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *19*(10), RC1.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is Belief in Free Will a Cultural Universal? *Mind & Language*, *25*(3), 346–358.

Schepers, R. J., Johanek, L. M., Hartke, T. V., Shim, B., Borzan, J., Meyer, R. A., & Ringkamp, M. (2008). A subpopulation of A-delta nociceptors in monkey is vigorously activated by cowhage spicules. In *Soc Neurosci Abstr* (Vol. 34).

Schieber, M. H., & Hibbard, L. S. (1993). How somatotopic is the motor cortex hand area? *Science*, *261*(5120), 489–492. doi:10.1126/science.8332915

Schlaghecken, F., & Eimer, M. (2004). Masked prime stimuli can bias" free" choices between response alternatives. *Psychonomic Bulletin and Review*, *11*, 463–468.

Schubotz, R. I., Friederici, A. D., & Yves von Cramon, D. (2000). Time Perception and Motor Timing: A Common Cortical and Subcortical Basis Revealed by fMRI. *NeuroImage*, *11*(1), 1–12. doi:10.1006/nimg.1999.0514

Schütz, A. C., & Morrone, M. C. (2010). Compression of time during smooth pursuit eye movements. *Vision research*, *50*(24), 2702–2713.

Schüür, F., & Haggard, P. (2011). What are self-generated actions? *Consciousness and Cognition*, *20*(4), 1697–1704.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the

cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154–167. doi:10.1038/nrn2994

Shadlen, M. N., & Gold, J. I. (2004). The neurophysiology of decision-making as a window on cognition. *The Cognitive Neurosciences*, 1229–1241.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*(2), 628 –633.

Sheather, S. (2004). Density Estimation. *Statistical Science*, *19*(4), 588–597.

Sherrington, C. S. (1906). *The integrative action of the nervous system*. Scribner.

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, *117*(11), 2341–2356. doi:10.1016/j.clinph.2006.04.025

Shima, K., Mushiake, H., Saito, N., & Tanji, J. (1996). Role for Cells in the Presupplementary Motor Area in Updating Motor Plans. *Proceedings of the National Academy of Sciences*, *93*(16), 8694–8698.

Shimazaki, H., & Shinomoto, S. (2010). Kernel bandwidth optimization in spike rate estimation. *Journal of computational neuroscience*, *29*(1), 171–182.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*(1), 174.

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. *Nat Neurosci*, *7*(1), 80–84. doi:10.1038/nn1160

Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. Retrieved from http://psycnet.apa.org/psycinfo/1939-00056-000

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946.

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2008). Movement-related potentials in the Go/NoGo task: The P3 reflects both cognitive and motor inhibition. *Clinical Neurophysiology*, *119*(3), 704–714. doi:10.1016/j.clinph.2007.11.042

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. doi:10.1016/j.tins.2004.01.006

Snyder, H. R., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition*, *116*(2), 155–167. doi:10.1016/j.cognition.2010.04.007

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, *11*(5), 543–545.

Spence, S. (2009). *The actor's brain: Exploring the cognitive neuroscience of free will*. Oxford University Press.

Stadler, W., Schubotz, R. I., Von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and memorizing observed action: Differential premotor cortex involvement. *Human Brain Mapping*, *32*(5), 677–687. doi:10.1002/hbm.20949

Staub, E., Tursky, B., & Schwartz, G. E. (1971). Self-control and predictability: Their effects on reactions to aversive stimulation. *Journal of Personality and Social Psychology*, *18*(2), 157.

Stroop, J. R. (1935). Studies of interference in verbal reactions. *Journal of Experimental Psychology*, *18*(6), 3–662.

Swann, N. C., Cai, W., Conner, C. R., Pieters, T. A., Claffey, M. P., George, J. S., … Tandon, N. (2011). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: Electrophysiological responses and functional and structural connectivity. *Neuroimage*.

Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *NeuroImage*, *56*(3), 1655–1665. doi:10.1016/j.neuroimage.2011.02.070

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324.

Tanji, J., & Hoshi, E. (2001). Behavioral planning in the prefrontal cortex. *Current Opinion in Neurobiology*, *11*(2), 164–170. doi:10.1016/S0959-4388(00)00192-6

Tanji, J., & Hoshi, E. (2008). Role of the Lateral Prefrontal Cortex in Executive Behavioral Control. *Physiological Reviews*, *88*(1), 37 –57. doi:10.1152/physrev.00014.2007

Tecce, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychological Bulletin*, *77*(2), 73–108.

Thaler, D., Chen, Y. C., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *102*(3), 445–460.

Thomas, E., & Brown, I. (1974). Time perception and the filled-duration illusion. *Attention, Perception, & Psychophysics*, *16*(3), 449–458. doi:10.3758/BF03198571

Thomas, S. (1894). *Summa theologica* (Vol. 6). Typographia Forzani et S.

Thut, G., Hauert, C.-A., Viviani, P., Morand, S., Spinelli, L., Blanke, O., … Michel, C. (2000). Internally driven vs. externally cued movement selection: a study on the timing of brain activity. *Cognitive Brain Research*, *9*(3), 261–269. doi:10.1016/S0926-6410(00)00004-5

Tolkien, J. R. R. (1954). The Lord of the Rings. 3 vols. *London: George Allen & Unwin*, *55*(1), 8.

Tomassini, V., Jbabdi, S., Klein, J. C., Behrens, T. E. J., Pozzilli, C., Matthews, P. M., … Johansen-Berg, H. (2007). Diffusion-Weighted Imaging Tractography-Based Parcellation of the Human Lateral Premotor Cortex Identifies Dorsal and Ventral Subregions with Anatomical and Functional Specializations. *The Journal of Neuroscience*, *27*(38), 10259–10269. doi:10.1523/JNEUROSCI.2144-07.2007

Treleaven-Hassard, S., Gold, J., Bellman, S., Schweda, A., Ciorciari, J., Critchley, C., & Varan, D. (2010). Using the P3a to gauge automatic attention to interactive television advertising. *Journal of Economic Psychology*, *31*(5), 777–784. doi:10.1016/j.joep.2010.03.007

Trevena, J., & Miller, J. (2009). Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition*.

Uematsu, S., Lesser, R., Fisher, R. S., Gordon, B., Hara, K., Krauss, G. L., … Webber, R. W. (1992). Motor and sensory cortex in humans: topography studied with chronic subdural stimulation. *Neurosurgery*, *31*(1), 59.

Vallesi, A., & Stuss, D. T. (2010). Excessive sub-threshold motor preparation for non-target stimuli in normal aging. *NeuroImage*, *50*(3), 1251–1257. doi:10.1016/j.neuroimage.2010.01.022

Van Eimeren, T., Wolbers, T., Münchau, A., Büchel, C., Weiller, C., & Siebner, H. R. (2006). Implementation of visuospatial cues in response selection. *NeuroImage*, *29*(1), 286–294.

Van Gaal, S., Ridderinkhof, K. R., Van den Wildenberg, W. P. M., & Lamme, V. A. F. (2009). Dissociating consciousness from inhibitory control: Evidence for unconsciously triggered response inhibition in the stop-signal task. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1129–1139. doi:10.1037/a0013551

Velmans, M. (2002). How could conscious experiences affect brains? *Journal of Consciousness Studies*, *9*(11), 3–29.

Verbruggen, F., & Logan, G. D. (2008a). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424. doi:10.1016/j.tics.2008.07.005

Verbruggen, F., & Logan, G. D. (2008b). Automatic and controlled response inhibition: Associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General*, *137*(4), 649–672. doi:10.1037/a0013170

Vogt, B. A., Nimchinsky, E. A., Vogt, L. J., & Hof, P. R. (1995). Human cingulate cortex: Surface features, flat maps, and cytoarchitecture. *The Journal of Comparative Neurology*, *359*(3), 490–506. doi:10.1002/cne.903590310

Vohs, K. D., & Schmeichel, B. J. (2003). Self-regulation and the extended now: Controlling the self alters the subjective experience of time. *Journal of Personality and Social Psychology*, *85*(2), 217–230.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will. *Psychological Science*, *19*(1), 49.

Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, *11*(2), 147–162.

Walsh, E., Kühn, S., Brass, M., Wenke, D., & Haggard, P. (2010). EEG activations during intentional inhibition of voluntary action: an electrophysiological correlate of self-control? *Neuropsychologia*, *48*(2), 619–626.

Walter, W. G., Cooper, R., Aldrige, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.

Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental brain research*, *162*(3), 346–356.

Watanabe, K., Igaki, S., & Funahashi, S. (2006). Contributions of prefrontal cue-, delay-, and response-period activity to the decision process of saccade direction in a free-choice ODR task. *Neural Networks*, *19*(8), 1203–1222. doi:10.1016/j.neunet.2006.05.033

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics*, *33*(2), 113–120.

Wegner, D. (2002). *The Illusion of Conscious Will (Cambridge*. MA.

Wegner, D., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480–492.

Wenke, D., Waszak, F., & Haggard, P. (2009). Action selection and action awareness. *Psychological Research*, *73*(4), 602–612. doi:10.1007/s00426-009-0240-4

Wittmann, M., Von Steinbüchel, N., & Szelag, E. (2001). Hemispheric specialisation for self-paced motor sequences. *Brain research. Cognitive brain research*, *10*(3), 341–344.

Wundt, W. M. (1909). *Über reine und angewandte Psychologie*.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science*, 182–186.

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, *414*(6861), 302–305. doi:10.1038/35104551

Yarrow, K., & Rothwell, J. C. (2003). Manual Chronostasis: Tactile Perception Precedes Physical Contact. *Current Biology*, *13*(13), 1134–1139. doi:10.1016/S0960-9822(03)00413-5

Yarrow, K., Whiteley, L., Rothwell, J. C., & Haggard, P. (2006). Spatial consequences of bridging the saccadic gap. *Vision Research*, *46*(4), 545–555. doi:10.1016/j.visres.2005.04.019

Zhang, J., Hughes, L. E., & Rowe, J. B. (2012). Selection and inhibition mechanisms for human voluntary action decisions. *NeuroImage*, *63*(1), 392–402. doi:10.1016/j.neuroimage.2012.06.058