Volition: From self-control to agency

Margaret T. Lynn

Promotor: Prof. dr. Marcel Brass

Proefschrift ingediend tot het behalen van de academische graad
van Doctor in de Psychologie

2014

VOLITION: FROM SELF-CONTROL TO AGENCY

Margaret T. Lynn

UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Volition: From self-control to agency

Margaret T. Lynn

Promotor: Prof. dr. Marcel Brass

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Psychologie

2014

Guidance Committee

Prof. dr. Marcel Brass (promotor)

Department of Experimental Psychology, Ghent University

Prof. dr. Wim Fias

Department of Experimental Psychology, Ghent University

Prof. dr. Roeljan Wiersema

Department of Experimental Clinical and Health Psychology, Ghent University

Prof. dr. Henk Aarts

Department of Psychology, Utrecht University

ACKNOWLEDGEMENTS

I am deeply grateful to have had so many wonderful people supporting me throughout this period of my life. Words could never do you justice, so I choose brevity, and trust you know how much you mean to me. This work is dedicated…

To Marcel, for challenging me and providing me with the opportunity to grow. I'm honored to have worked with you.
To my guidance committee, for asking the hard/right questions.
To Ruth, Ezequiel, and Henk, for teaching me what makes a great researcher.
To Jelle, for being a fantastic office mate and collaborator.
To my colleagues (esp. Elisah, David, Nico, Egbert, Marlies, Michel, Elena, Valerie, Jim, Leonie, Charlotte, Eliana, Margarita, and Jasmina), for giving me hope. And to Anand for being my stand-in brother.
To my beloved PEGs, for giving me a forever tribe.
To my San Francisco pack (including those now elsewhere), for utter joy.
To everyone I've missed while across the ocean (esp. Shanna, Mani, Anne, Stu, Guzel, Pam, Sheila, and Sandy) – my heart bursts thinking of the reunions. And to those on this side of the pond (Melissa, Kevin, Emre, and Scott), for the much needed R&R.
To Naomi, for so much laughter and commiseration.
To Patti, for her depth, modesty, steadfastness, and humor.
To Doro, for helping me through heartache and encouraging my creative side.
To my parents, for always fostering my love of knowledge, and for giving me the world at such a young age.
To Jake, for believing in me. I'm proud to be your sister.
Most of all, to Paul, for changing everything about my life in the best possible way.

Maggie

October 2014

CONTENTS

Imagine that, deeply immersed in thought, you inadvertently walk home on autopilot when you meant to go to the store instead. You have no memory of the route, how much time it took, nor even who or what you might have passed along the way. If asked, you would hardly classify this as a voluntary act, and yet to an outside observer it is likely to appear that you have purposefully chosen your path. What constitutes the feeling of volition (i.e., will; the faculty of choice and determination by which we control ourselves or our environment) is often an entirely internal experience, with descriptors such as *willing, intending, desiring,* and *being inclined.* It is the subtle difference between a blink and a wink (Morsella, Molapour, & Lynn, 2013). Undoubtedly, volition is fundamental to what it means to be human; we diverge from other species due to this specific power of deliberation. While the behavior of other species is largely determined by environmental triggers, our ability to control behavior endogenously grants us *freedom from immediacy* (Shadlen & Gold, 2004).

Problems inherent in volition as a scientific concept

Lacking omniscient access to the internal states of others, scientists have been faced with the struggle to define what constitutes a volitional act and how it may be measured. Phenomenologically, a volitional act is characterized by its first-person ontology: An act is volitional if I feel that I have acted as intended. This conception of volition is reflected in the introspective tradition of the early days of empirical psychology research (e.g., Harleß, 1861; James, 1890; Lotze, 1852). However, since the rise of behaviorism, introspection has been dismissed for its subjectivity and largely abandoned as a method for psychological research. Alternatively, volitional action can be defined by comparing it with stimulus-driven action (e.g., Herwig, Prinz, & Waszak, 2007; Herwig & Waszak, 2009; Waszak et al., 2005). This definition is more fruitful for empirical research, as it allows for the comparison of actions that differ in terms of the locus of decision yet are similar in terms of movement kinematics. This conception contains several problems, however: First, it limits the study of volition to rather simple movements such as button presses. Second, it is quite controversial to what extent internal factors can be traced experimentally (Nachev & Husain, 2010). Finally, there is an increasing consensus that, in everyday experience, actions are never exclusively volitional or stimulus-driven, but rather exist along a continuum of “reflexiveness” (see

Krieghoff, Waszak, Prinz, & Brass, 2011). Hence, the difference between endogenous and exogenous causes may blur, particularly when similar actions are executed repetitively, as in most psychological experiments.

Beyond denotation, another challenge when studying volition is the design of suitable experimental paradigms. Volitional behavior is characterized by its reliance on endogenous rather than exogenous determinants (Haggard, 2008). Thus, any instruction from the experimenter that specifies how the participant should behave ultimately calls into question to what extent subsequent behavior should still be categorized as volitional. The catch is that a certain level of instruction is simply necessary to obtain meaningful data that can be quantified and compared across participants. The most prevalent strategy in the literature has been to give incomplete instructions, in which certain movement parameters (e.g. the selection or the timing of a movement) are left free to be generated by the participant (Haggard & Eimer, 1999; Krieghoff, Brass, Prinz, & Waszak, 2009; Libet, Gleason, Wright, & Pearl, 1983). Although such "gaps" in task instruction certainly draw upon internal generation, they confine volition to very specific aspects of the decision process. A final cardinal problem in the study of volition concerns motivational aspects. Volitional behavior is typically directed toward a desired outcome (see Haggard, 2008). In other words, we act or suppress action because we want to bring about some (short-term or long-term) change in our environment. Thus, paradigms intended to inform volition should create experimental contexts in which behavior is not only endogenously generated but also motivated (rather than, e.g., choosing between left-hand and right-hand button presses). Altogether, these considerations illustrate the obstacles inherent in studying volition as a scientific construct; one is faced with an inevitable tradeoff between internal and ecological validity, an issue that will be revisited a number of times in this thesis.

Studying volition

In spite of these challenges, several effective strategies have been used to work around the aforementioned problems and measure volition in a scientifically acceptable and ecologically valid way. In the following, I will briefly describe three such strategies that have been particularly informative. The first is to employ indirect or implicit

measures of volitional processes rather than to rely on introspection. For instance, the perceptual attraction between a voluntary action and its sensory outcome has been validated as a measure of the sense of agency (i.e. the experience of causing the action outcome, see Moore & Obhi, 2012 and later sections). In this sense, implicit measures are directly observable and quantifiable variables that stand in some relation to the unobservable volitional process of interest. A second strategy is to employ functional brain imaging while subjects engage in volitional behavior. Neuroimaging permits measurement of an observable correlate of volitional processes in the human brain, and as such is another means of materializing and quantifying certain volitional acts. Research along these lines has revealed that endogenous control of behavior relies on the medial frontal cortex, whereas the contextual guidance of behavior on the basis of external information relies more strongly on the lateral part of the frontal lobe (Amodio & Frith, 2006; Goldberg, 1985; Passingham, Bengtsson, & Lau, 2010; Rushworth, Walton, Kennerley, & Bannerman, 2004)[1]. Moreover, recent neuroimaging studies have begun to decompose the broad concept of volition into distinct components on the basis of their neural signature (Brass & Haggard, 2008; Brass, Lynn, Demanet, & Rigoni, 2013; Haggard, 2008; Krieghoff et al., 2009). However, despite its great appeal, a caveat needs to be kept in mind: this method often relies crucially on reverse inference; i.e., the inference of a psychological process based on the activity of a particular brain region, which is deductively invalid and potentially misleading (Poldrack, 2006). A final strategy is to study neuropsychiatric patients for whom particular components of volitional control have gone awry (e.g., Brandt, Lynn, Obst, Brass, & Münchau, 2014). For instance, anarchic hand syndrome, which can result from brain lesions to the pre-supplementary motor area, is characterized by an inability to suppress an action tendency that is induced by an external stimulus despite having the subjective experience of doing so (e.g., Della Sala, Marchetti, & Spinnler, 1991; Kritikos, Breen, & Mattingley, 2005; Pacherie, 2007). Such findings are highly informative about the phenomenology, the neural basis, and the functional mechanisms that give rise to volitional behavior and experiences. Importantly, each of these strategies has distinct

[1] In line with this conceptual distinction, the medial frontal cortex has extensive connections with areas that are involved in processing internal states, whereas the lateral frontal cortex has stronger connections with sensory areas that infer details of the external world (see Passingham et al., 2010; Passingham & Wise, 2012).

strengths and limitations, and a satisfying science of volition requires integration across methods.

Other lines of research have focused less on the architecture of volition, but rather on the identification of factors that influence the ability or motivation to exert volitional control. For example, a host of studies have examined the idea that volition relies on a limited self-regulatory resource, and that exhaustion thereof (via effortful acts of self-control) results in an ultimate breakdown of volitional control on subsequent tasks (*ego depletion*, see Baumeister, Bratslavsky, Muraven, & Tice, 1998). Other studies have focused on the influence of high-level beliefs on behavioral control. In particular, beliefs about the concept of volition (or *free will*) have received increasing attention; it has been shown that undermining the strength of these beliefs has a substantial impact on behavior in social contexts (e.g., Baumeister, Masicampo, & Dewall, 2009; Vohs & Schooler, 2008) and even on basic motor control processes (Rigoni, Kühn, Sartori, & Brass, 2011), although the underlying mechanisms remain to be understood (see later sections).

In summary, various means of studying volition have been developed, and can be broadly categorized into (i) research that attempts to measure particular components of volition, and (ii) research that seeks to identify factors that influence the exertion of volitional control. This distinction also applies to the structure of the present thesis. In the first part, I will attempt to develop a novel and ecologically valid way of measuring volition in a particular domain, namely the control of behavioral impulses. Thereafter, in the second part of my thesis, my goal is to identify factors that impact the ability to exert volitional control.

Part I: Volition as impulse control

One way volitional behavior can be defined is by its capacity to be suppressed (Passingham, 1993). We experience a wide variety of impulses each day, some stimulus-driven, some internally generated. Yet these impulses may not always be concordant with long-term goals or societal norms. Accordingly, the ability to control and override unwanted behavioral impulses is a hallmark of adaptive behavior in complex social systems. Such instances of self-control have been studied in diverse research traditions.

Evidence from social psychology

Social psychological perspectives on impulse control tend to be concerned with the situational or dispositional factors that determine one's self-regulatory abilities, and how this faculty might be perturbed. Metcalfe and Mischel (1999) originally proposed a dual-systems framework for the exertion of self-control, in which the *hot* emotional system responds to impulsive motivational triggers, while the *cool* cognitive system operates in a reflective and deliberative manner to override the hot system when necessary. In this sense, the hot system is directly under stimulus control, while the cool system incorporates prior knowledge to achieve long-term goals. Other work has focused on delineating psychological strategies that people apply in order to resist impulses such as distraction, forming implementation intentions, or re-appraisal (e.g., Fischbach, Friedman, & Kruglanski, 2003; Gollwitzer & Brandstätter, 1997; Mischel & Baker, 1975), and designing methods to enhance the ability for impulse control (e.g., Webb & Sheeran, 2003). Self-control has further been quantified at the trait level, and is predictive of long-term outcomes such as job success, health, self-esteem, and prosocial behavior (Mischel et al., 2011; Tangney, Baumeister, & Boone, 2004). Not only has it been shown that individuals high in trait self-control perform better on tests of inhibition (Schmeichel & Zell, 2007), but when controlling for individual differences in the frequency of impulsive tendencies, Friese and Hofmann (2009) found that the expression of behavioral urges occurs much more readily in people low in trait self-control. Overall, the social psychological approach to impulse control is characterized by its diversity of paradigms in which impulsive behavior is assessed, ranging from performance at simple laboratory tasks (e.g., the Stroop task) to complex real life behaviors (e.g., dieting).

Evidence from cognitive psychology and neuroscience

In cognitive psychology and neuroscience, a quite different approach has been employed when studying the capacity to control impulsive behavior. Here, research is concerned with the delineation of the core functional and neural mechanisms that enable the stopping of a prepotent action tendency under strictly controlled experimental conditions. The most commonly used tasks to assess this ability are the

stop signal task and the go/nogo task (see Verbruggen & Logan, 2009 for a review).[2] In both tasks, participants are required to perform speeded responses to the identity of a frequently presented go stimulus, inducing a strong response tendency or "readiness to act." Occasionally, however, a stop or nogo signal is presented, instructing participants to withhold the prepotent response. Nogo signals are typically presented as target stimuli *instead* of the go signal and the dependent measure the error rate on nogo trials (i.e., the success of stopping). Stop signals, on the other hand, are typically presented *after* the go signal (e.g., as a sound), thus requiring a cancellation of a response at a point in time in which it is already prepared or initiated. This conflict between going and stopping is typically conceived as a race between a go process and a stop process, in which the process that is completed first determines the outcome (see Logan & Cowan, 1984; Verbruggen & Logan, 2008). By systematically manipulating the time lag between the go signal and the stop signal, and taking into account a participant's reaction time on go trials, it is possible to estimate the speed of an individual's stopping process, referred to as the stop-signal reaction time (SSRT). This measure is commonly employed as an index of individual stopping ability.[3]

Over the past decade, myriad neuroscience studies have employed variants of the stop signal and go/nogo task to reveal the neural mechanisms that are involved in the suppression of action tendencies. Multi-method evidence suggests that stopping in response to the occurrence of nogo or stop signals relies critically on the integrity of the right inferior frontal gyrus (rIFG), a small part of the lateral prefrontal cortex (PFC) that lies rostral to the precentral sulcus and ventral to the inferior frontal sulcus (see Aron, Robbins, & Poldrack, 2014 for a recent review). This region is not only reliably activated during stopping (e.g., Garavan et al., 2002; Garavan, Ross, & Stein, 1999; Konishi et al., 1999; Konishi et al., 1998), but disruption of its integrity – either via natural lesions or by means of transcranial magnetic stimulation (TMS) – results in a substantial increase in SSRT, even when controlling for damage to other PFC sub-regions (Aron et al., 2003; Chambers et al., 2006; Verbruggen & Aron, 2010). Notably, despite copious evidence

[2] Note that there are numerous other tasks that may also rely on inhibitory control, such as the antisaccade task (see Hutton & Ettinger, 2006), but these tasks also involve a number of non-inhibitory executive functions and are therefore not discussed in detail here.

[3] Note that the exact estimation of SSRT is still somewhat controversial, as it might be sensitive to strategic adjustments (see Boehler, Hopf, Stoppel, & Krebs, 2012; Verbruggen, Chambers, & Logan, 2013).

that the rIFG is essential for the ability to stop a prepotent tendency in response to a stop or nogo signal, it has been proposed that its role might be better characterized as an attentional monitor for salient external events (e.g., Erika-Florence, Leech, & Hampshire, 2014; Hampshire et al., 2010). However, recent evidence that signals from the rIFG directly suppress the excitability of the primary motor cortex (Neubert et al., 2010) seems hard to reconcile with an exclusively attentional account.

Beyond the rIFG, the pre-supplementary motor area (preSMA) is another cortical region that has been implicated in behavioral inhibition (see Chambers, Garavan, & Bellgrove, 2009; Nachev, Kennard, & Husain, 2008). The preSMA is located in the posterior medial frontal wall, rostral to the supplementary motor area and dorsal to the cingulate gyrus. It is connected with the rIFG structurally and functionally (Aron, 2007; Johansen-Berg et al., 2004) and also exhibits reliable activity during stopping (e.g., Boehler et al., 2010; Duann, Ide, Luo, & Li, 2009; Swann et al., 2012). Moreover, as mentioned above, unilateral lesions to the preSMA can lead to *anarchic hand syndrome*. Patients with anarchic hand syndrome are unable to inhibit stimulus-induced action tendencies with the hand contralateral to the lesion, even though they report having an intention to do so (Della Sala, Marchetti, & Spinnler, 1991; Kritikos, Breen, & Mattingley, 2005; Pacherie, 2007). Considerable controversy has evolved regarding the differential contributions of the preSMA and the rIFG to stopping (e.g., Chevrier, Noseworthy, & Schachar, 2007; Chikazoe et al., 2009; Li, Huang, Constable, & Sinha, 2006; Neubert et al., 2010; Sharp et al., 2010). In this regard it is quite noteworthy that, whereas the rIFG seems to be recruited rather specifically for response inhibition, the preSMA appears to contribute to cognitive control more generally. For instance, activation of the preSMA is often found in the Stroop task or during task switching (Niendam et al., 2012), and it is not clear to what extent these tasks require inhibitory control. Moreover, the preSMA also contributes to the planning and generation of self-initiated movements. Accordingly, it has been proposed that the pre-SMA may be critical for adjusting the degree of action readiness (Cunnington, Windischberger, & Moser, 2005; Forstmann et al., 2008) and implementation of action decisions that can yield facilitation or suppression of motor output.

Current models of response inhibition assume that the rIFG and the preSMA implement behavioral inhibition via interactions with subcortical regions. In particular

the subthalamic nucleus (STN), a small basal ganglia input nucleus, has been ascribed a prominent role. It receives direct projections from both the rIFG and the preSMA and is functionally coupled with these regions on successful stop trials (Aron, 2007; Inase et al., 1999). Models typically also distinguish between a *hyperdirect pathway* that involves only direct projections from the cortex to the STN and leads to fast and global motor suppression (i.e., suppression of all motor output) and an *indirect pathway* that also involves the striatum and the pallidum and allows for less instantaneous, yet effector-specific, inhibition (see Aron et al., 2014; Aron, 2011 for a detailed overview).

Limitations of previous research on impulse control

It should be evident from the previous section that the social psychological and cognitive approaches toward the study of impulse control differ substantially from one another, and entail complementary strengths and limitations in revealing the mechanisms of impulse control. The social approach greatly benefits from its ecological validity, as it examines self-regulatory behaviors in realistic scenarios. Moreover, it allows for the investigation of the impact of subjective variables on self-control, which are typically neglected in cognitive studies. That being said, in the more naturalistic settings of social psychological studies it is very difficult to reveal the precise mechanisms that contribute to self-regulation, as the tasks often permit multiple strategies to be employed. Conversely, the cognitive approach offers maximal experimental control, which is suitable for the delineation of specific componential processes and the characterization of the involved neural systems. Yet classical cognitive paradigms, such as the SST or the go/nogo task fall short in terms of ecological validity. As the decision to implement or inhibit an action tendency is fully determined by external stimuli in these tasks, it is questionable to what extent conclusions may be generalized to impulse control in the real world, where self-control must typically originate endogenously.

In recent years, efforts have been undertaken to bridge the gap between experimental scenarios of response inhibition and impulse control in everyday life. One approach has been to extend classical response inhibition paradigms in order to increase their reliance on endogenous and motivational processes. As such, the concept of proactive inhibition has been introduced (e.g., Cai, Oldenkamp, & Aron, 2011;

Chikazoe et al., 2009; Jaffard et al., 2008; Jahfari, Stinear, Claffey, Verbruggen, & Aron, 2010; Wessel & Aron, 2014). Proactive inhibition is assessed in a modified version of the SST by giving participants advance information about the probability of a stop signal in the current block. It is then measured how this knowledge affects the proactive preparation to stop by comparing blocks without stop signals (never stop; NS) with blocks that contain a moderate number of stop signals (maybe stop; MS). Behaviorally, it has been documented that anticipating the possibility of a stop leads to elongated reaction time on MS go trials, indicating strategic adjustments (Verbruggen & Logan, 2009). Neuroimaging studies have moreover revealed greater rIFG activity on MS go trials compared with NS go trials, with the amount of IFG activity predicting the amount of slowing (Jahfari et al., 2010). These findings highlight that the aforementioned network for response inhibition is not only recruited reactively by the occurrence of a stop or nogo signal, but also proactively by the anticipation thereof. Another interesting approach has been the inclusion of motivational incentives into response inhibition paradigm. For instance, Boehler et al. (2012) demonstrated that the SSRT is reduced when a stop signal indicates that a monetary reward can be gained via successful stopping, presumably due to a stronger recruitment of brain areas implicated in stopping (Boehler et al., 2014). Although both of these extensions of classical response inhibition paradigms have yielded novel and valuable insights, some of the aforementioned limitations of cognitive paradigms still apply. Most importantly, the decision to act or inhibit in these paradigms is still entirely stimulus-driven, limiting its validity as a model of self-control in everyday life.

As an alternative approach toward the study of the neural basis of self-control, the concept of *intentional inhibition* has been introduced to the literature. Intentional inhibition describes the suppression of prepotent behavior on the basis of endogenous decisions. Accordingly, this concept of inhibition is more closely related to social psychological concepts of self-control. To date, only a few studies have examined this capacity, and the paradigms that have been employed are quite diverse. In the first study by Brass & Haggard (2007), participants were required to press a button at moment of their choosing (action trials), and to occasionally prepare a button press but cancel it at the last possible moment (inhibition trials). Contrasting brain activity on inhibition trials with action trials revealed no activation in the areas previously

implicated in inhibition in the SST. Instead, inhibition-related activity was found in the dorsal frontal median cortex (dFMC), located in the anterior medial frontal wall rostral and dorsal to the anterior cingulate, along with the anterior insula and the inferior parietal cortex. This finding was interpreted as reflecting the existence of two inhibition systems for externally guided and intentional inhibition respectively[4]. Importantly, several follow-up studies could replicate the involvement of the dFMC in intentional inhibition using disparate experimental procedures such as omitting a prepotent button press that could prevent an aversive glass-breaking sound and a monetary loss (Kühn, Haggard, & Brass, 2009), or the voluntary suppression of negative emotions (Kühn, Haggard, & Brass, 2014).[5] Based on these findings, it was proposed that the dFMC constitutes a "veto area" that serves the suppression of impulsive behavior on the basis of endogenous decisions (Brass & Haggard, 2007, 2008).

However, although intentional inhibition studies certainly address some of the shortcomings of previous response inhibition research, the transfer of these findings to realistic instances of self-control remains difficult. Most importantly, although the decisions to inhibit were generated endogenously, they were often also rather arbitrary. Neither option had a genuine motivational valence to the participant (e.g., choosing between pressing a button or refraining from pressing it), in striking contrast to self-control in real life, which is characterized by the presence of strong urges to act that conflict with an individual's goal or social norms. Accordingly, it remains to be determined to what extent these findings would generalize to contexts in which inhibition of a highly prepotent response is required. Moreover, the mechanism through which intentional inhibition is achieved remains elusive. Does it reflect motor suppression in a strict sense, or rather higher-order self-regulatory processes? The paradigms that have been employed thus far suggest that it reflects a very general mechanism that might also involve non-motor processes, but further characterization seems necessary.

[4] This distinction directly corresponds to the aforementioned idea that endogenous and exogenous control rely more strongly on the medial and lateral frontal cortex respectively (Goldberg, 1985; Passingham et al., 2010).

[5] Note, however, that the precise anatomical location of the dFMC peak coordinates varied considerably across these studies.

In summary, several novel and well-motivated approaches have been employed to relate the neuroscience of response inhibition more closely to self-control in real life. However, despite valuable advances from original designs, fundamental problems remain. In particular, paradigms either provide participants with external information that determines the decision to act or inhibit, or they employ rather arbitrary decision scenarios, absent behavioral urges. Both of these aspects limit the ecological validity of paradigms as a model of impulse control. Accordingly, one central aim of the present thesis will be the development of a novel experimental paradigm that may help to measure behavioral inhibition in a more valid and yet experimentally controlled manner.

Part II: Determinants of volitional engagement

The second part of my thesis shifts the focus from measuring and characterizing volition itself towards delineating factors that impact the ability and/or motivation to exert volitional control. This question has been addressed in different areas of social and cognitive psychology that will be briefly reviewed in the following.

Self-regulatory abilities

A controversial debate in the social psychological literature has been centered around the question of whether volition relies on a limited self-regulatory resource (*willpower*; Baumeister, 2003; Baumeister, Bratslavsky, Muraven, & Tice, 1998), a capacity for volitional action capable of being depleted following repeated exertions (*ego depletion*; Baumeister et al., 1998). Tasks requiring willpower include self-control, decision-making, complex problem solving, and conflict resolution. From this perspective there is not one task that depletes the human will, but rather a number of tasks that draw more or less on this resource. In a series of behavioral studies, Baumeister and colleagues showed that differing tasks presumed to rely on willpower systematically interfered with each other (Baumeister, 2003; Muraven & Baumeister, 2000). However, despite the diversity of studies investigating ego depletion, to date little is known about what might constitute a biological substrate of willpower. Gailliot & Baumeister (2007) proposed that consumption of glucose might underlie the

depletion of self-regulatory capacity. However, this proposal has been largely falsified (Beedie & Lane, 2011) and alternative proposals are lacking.

A related line of research has focused on stable inter-individual differences in self-regulatory abilities rather than on its situational depletion (i.e., on *trait* differences in self-regulation rather than *state* differences). Ample evidence suggests that stable traits, reflecting self-control abilities, build up early in development and are predictive of various long-term outcomes. The most striking evidence comes from the famous delay of gratification experiments (Mischel, 1974), in which preschoolers could chose between an immediately available but small reward, and a larger reward with a temporal delay. The researchers found marked differences in the ability to override the impulse to give in the temptation of the immediate reward to obtain the favorable long-term outcome. Importantly, longitudinal assessments revealed that the ability to delay gratification at preschool age were predictive of participants' success later in life, determined by factors such as employment and standardized testing performance, illustrating both the stability of trait self-control and the ecological validity of its measurement (see Mischel et al., 2011 for a contemporary review). However, as in the case of ego depletion, little if anything is known with regard to the mechanisms that underlie different levels of self-regulatory ability. Accordingly, one goal of the second part of this thesis will be to examine the neural underpinnings of differences in self-regulatory abilities, both at the state level and at the trait level.

High-level beliefs

High-level beliefs regarding volition (or free will) have also been related to the magnitude of volitional engagement. The sensation of having free control over one's actions is an undeniably ubiquitous feature of human experience. Perception of personal control is further considered to be intrinsic, biologically necessary, and protective against environmental stressors (Leotti, Iyengar, & Ochsner, 2010). A recent line of research has examined the effects of undermining free will beliefs (FWBs) experimentally. Typically, one group of participants is required to read an essay that promotes a determinist perspective by questioning the concept of humans as self-determined intentional agents (e.g., "all behavior is determined by our genes"), whereas another group is required to read a control text that contains general

statements about consciousness. A number of studies have shown that these manipulations can not only reduce the strength of participants' self-reported FWBs, but also lead to a degradation of behavioral control in social contexts. For instance, Vohs and Schooler (2008) found that participants whose FWBs were weakened paid themselves a statistically improbable amount of money for performance on a problem-solving task, and more frequently permitted themselves to view answers when given the opportunity to cheat. In the same vein, undermining FWBs has been shown to increase aggressive behaviors and decrease pro-social behavior (Baumeister et al., 2009), promote mindless conformity (Alquist, Ainsworth, & Baumeister, 2013) and reduce counterfactual thinking (Baumeister, Crescioni, & Alquist, 2010). Altogether, these findings indicate that a stable belief in free will might be necessary to maintain the motivation that is necessary to exert effortful control over selfish impulses in favor of socially appropriate behavior that concurs with societal norms.

More recently, cognitive psychologists have revealed that the effects of free will manipulations are not restricted to behavior in social contexts, but seem to propagate to very basic and even unconscious stages of motor control. Rigoni et al. (2011) employed a similar belief manipulation and measured its effects by means of a Libet task, in which participants performed self-initiated button presses and were required to monitor the point in time at which they experienced the intention to move (see Libet et al., 1983). Electroencephalographic recordings of scalp currents prior to the movements revealed that participants whose FWBs had been perturbed exhibited a reduced amplitude of the readiness potential, an electrophysiological marker of endogenous motor preparation (Shibasaki & Hallett, 2006), highlighting that FWBs may affect even the most basic stages of action control.

Thus, a substantial body of evidence indicates that FWBs play an important role in behavioral control. Yet, despite these convergent findings, it is still poorly understood how weakening the strength of FWBs influences human behavior and how it can affect such a wide variety of control processes. Rigoni et al. (2011) speculated that reducing FWBs might perturb participants' *sense of agency* (SoA), i.e., their intrinsic experience of being in control of their actions, which may then, in turn, reduce the recruitment of intentional effort in action production and impulse control. However, although this

view is principally plausible, a conclusive link between FWBs and agency remains to be established and constitutes another goal of the present thesis.

Agency

The sense of agency refers to the experience of being in control of one's own actions and, consequently, of events in the external world. It is an intrinsic and ubiquitous experience that, when intact, rarely rises to the level of conscious awareness (see Chambon, Filevich, & Haggard, 2014; Haggard & Chambon, 2012; Haggard & Tsakiris, 2009; Morsella, Berger, & Krieger, 2010 for recent reviews). However, several clinical disorders are characterized by a perturbation of the SoA. For instance, patients with psychosis often report experiencing abnormal intentions and causality attributions (Jeannerod, 2009; Kapur, 2003). In the healthy population, deviating agentic experiences can be witnessed in instances such as losing control over a car that suddenly malfunctions.

Classical theoretical accounts of the SoA have proposed that it is based on a mental "comparator" that constantly predicts the sensory outcomes of ongoing actions and matches them with actual sensory input (Frith, Blakemore, & Wolpert, 2000; Wolpert, Ghahramani, & Jordan, 1995). According to this view, we have a continuous experience of being in control as long as predictions and sensory input correspond, but in cases of a strong mismatch the default experience of "I did it" will suddenly become disrupted. More recent models have extended the comparator metaphor and typically distinguish between a nonconceptual or implicit component and a conceptual or explicit component. The former is often referred to as a pre-reflective and intrinsic feeling of being in control, and as a fluid experience of causing one's own actions and their outcomes. The latter is often referred to as reflective or explicit judgment, or attribution of authorship based on deliberate thought processes (e.g., Synofzik, Vosgerau, & Newen, 2008).

This conceptual distinction between explicit and implicit components of agency also reflects the types of paradigms that have been employed to measure the SoA. Explicit agency components are typically measured in attribution tasks that require participants to identify the generator of an "ambiguous" event in their environment. For example, some tasks involve participants moving a joystick while monitoring

changes on a video screen; occasionally they are asked about their certainty that they have caused a particular change on the screen to appear. Research along these lines has identified several factors that contribute to explicit attributions of agency, among them the temporal proximity between an action and its outcome, or the consistency with prior action-outcome experiences (e.g., Sato & Yasuda, 2005). By contrast, implicit agency components are typically measured indirectly, most commonly with the so-called intentional binding (IB) task (see Moore & Obhi, 2012 for a review). In this task, participants perform temporal judgments about movements or about sensory events when these events occur in isolation (baseline) or in combination (agency), i.e., when the tone follows shortly after and is contingent with the action. IB denotes the observation that there is a systematic perceptual shift in the agency conditions. Compared to the baseline conditions, actions are perceived as occurring later, and sensory events as occurring earlier. In other words, the action and its outcome are bound together in time. Importantly, ample evidence suggests that IB is sensitive to the level of action intention, as it is reduced or even absent when movements are not self-initiated but rather guided by external stimuli or triggered by TMS (Engbert, Wohlschlager, & Haggard, 2008; Engbert, Wohlschlager, Thomas, & Haggard, 2007; Haggard, Clark, & Kalogeras, 2002), motivating its common use as a measure of implicit agency.

Overview of the chapters

As outlined above, my thesis is structured into two distinct lines of research. The first line deals with the neural basis of inhibitory self-regulation, with the goal of providing an ecologically valid way to investigate this capacity. To this end, in **Chapter 1**, I introduce a novel paradigm that assesses behavioral inhibition in the context of pain avoidance behavior. Participants received thermal pain stimulation to alternating inner-wrists and could terminate the stimulation via a button press with the non-stimulated hand. On some trials, the decision to act or inhibit was indicated by an external cue, whereas on other trials, participants could freely chose between both options. The advantage of this paradigm is that it induces a strong urge to act in each participant and on every trial. Response inhibition thus mirrors realistic instances of self-control more closely than in previous paradigms. Strikingly, under these motivationally salient

conditions, inhibition relies on joint activation of the brain networks previously implicated in externally-guided and intentional inhibition. This finding casts doubt on the original assumption that these networks serve independent functions and indicates that they might operate in concert in the context of genuine self-control. In **Chapter 2,** I integrate these and other recent neuroimaging findings into a novel theoretical account of the role of the dorsal frontomedian cortex in intentional inhibition. Based on a number of studies that have linked this region with non-motor and self-referential processes, I argue that the contribution of this region to self-regulation might be better characterized as higher-level disengagement from intentions and urges rather than as a veto signal that cancels motor plans in a strict sense.

In the second line of research, I identify factors that perturb the ability to exert volitional control. Having established the mechanisms involved in the inhibition of pain avoidance behavior, **Chapters 3 and 4** use an adapted version of the same paradigm. **Chapter 3** investigates the relation between inhibition-related brain activity and individual differences in self-regulatory abilities (both at the state level and at the trait level). Results indicate that a high trait level of self-regulation is associated with stronger activation of the inhibition network, as defined in the previous study, whereas no effects of the state manipulation were found. **Chapter 4** examines beliefs about volition and their impact on behavioral control by investigating whether weakening FWBs affects intentional engagement in self-control. Some evidence indicates that this might be the case. However, the effects are only found in a subset of the sample, i.e., in those participants who experience a sufficient amount of pain to necessitate self-control. **Chapter 5** seeks to reveal how FWBs are able to affect various behavioral parameters, ranging from the readiness potential to complex social behaviors. Specifically, it tests the hypothesis that weakening FWBs diminishes the sense of agency, which is largely confirmed by the data.

References

Alquist, J. L., Ainsworth, S. E., & Baumeister, R. F. (2013). Determined to conform: Disbelief in free will increases conformity. *Journal of Experimental Social Psychology*, *49*, 80–86.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–277.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist*, *13*, 214–228.

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biological Psychiatry*, *69*(12), e55–68.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, *6*(2), 115–6.

Aron, A. R., Robbins, T. W., & Poldrack, R. a. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–85.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252–1265.

Baumeister, R. F., Crescioni, a. W., & Alquist, J. L. (2010). Free Will as Advanced Action Control for Human Social Life and Culture. *Neuroethics*, *4*(1), 1–11.

Baumeister, R. F., Masicampo, E. J., & Dewall, C. N. (2009). Prosocial benefits of feeling free: disbelief in free will increases aggression and reduces helpfulness. *Personality & Social Psychology Bulletin*, *35*(2), 260–8.

Beedie, C. J., & Lane, A. M. (2011). The role of glucose in self-control: Another look at the evidence and an alternative conceptualization. *Personality and Social Psychology Review, 16*(2), 143 - 153.

Boehler, C. N., Appelbaum, L. G., Krebs, R. M., Hopf, J. M., & Woldorff, M. G. (2010). Pinning down response inhibition in the brain--conjunction analyses of the stop-signal task. *NeuroImage*, *52*(4), 1621–32.

Boehler, C. N., Hopf, J.-M., Stoppel, C. M., & Krebs, R. M. (2012). Motivating inhibition - reward prospect speeds up response cancellation. *Cognition*, *125*(3), 498–503.

Boehler, C. N., Schevernels, H., Hopf, J.-M., Stoppel, C. M., & Krebs, R. M. (2014). Reward prospect rapidly speeds up response inhibition via reactive control. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 593–609.

Brandt, V. C., Lynn, M. T., Obst, M., Brass, M., & Münchau, A. (2014). Visual feedback of own tics increases tic frequency in patients with Tourette's syndrome. *Cognitive Neuroscience*, October, 1–7.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *The Journal of Neuroscience*, *27*(34), 9141–5.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist*, *14*, 319–325.

Brass, M., Lynn, M. T., Demanet, J., & Rigoni, D. (2013). Imaging volition: what the brain can tell us about the will. *Experimental Brain Research, 229*(3), 301-312.

Cai, W., Oldenkamp, C. L., & Aron, A. R. (2011). A proactive mechanism for selective suppression of response tendencies. *The Journal of Neuroscience*, *31*, 5965–5969.

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., … Mattingley, J. B. (2006). Executive "brake failure" following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience*, *18*(3), 444–55.

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews*, *33*(5), 631–46.

Chambon, V., Filevich, E., & Haggard, P. (2014). What is the human sense of agency, and is it metacognitive? In S. M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 321–342). Berlin: Springer.

Chevrier, A. D., Noseworthy, M. D., & Schachar, R. (2007). Dissociation of response inhibition and performance monitoring in the stop signal task using event-related fMRI. *Human Brain Mapping*, *28*, 1347–1358.

Chikazoe, J., Jimura, K., Hirose, S., Yamashita, K., Miyashita, Y., & Konishi, S. (2009). Preparation to inhibit a response complements response inhibition during performance of a stop-signal task. *The Journal of Neuroscience*, *29*, 15870–15877.

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.

Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, *29*, 1113–1117.

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*, *29*(32), 10171–9.

Engbert, K., Wohlschlager, A., & Haggard, P. (2008). Who is causing what? The sense of agency is rational and efferent-triggered. *Cognition, 107*(2), 693-704.

Engbert, K., Wohlschlager, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology - Human Perception and Performance, 33*(6), 1261-1268.

Erika-Florence, M., Leech, R., & Hampshire, A. (2014). A functional network perspective on response inhibition and attentional control. *Nature Communications*, *5*(May), 4073.

Fischbach, A. Friedmann, R. S., & Kruglanski, A. W. (2003). Leading us not into temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology, 84,* 296 - 309.

Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2008). Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: a model-based approach. *The Journal of Neuroscience*, *28*, 9790–9796.

Friese, M., & Hofmann, W. (2009). Control me or I will control you: Impulses, trait self-control, and the guidance of behavior. *Journal of Research in Personality*, *43*(5), 795–805.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, *31,* 357–363.

Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: linking blood glucose to self-control. *Personality and Social Psychology Review*, *11*, 303–327.

Garavan, H., Ross, T. J., Murphy, K., Roche, A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection and correction. *NeuroImage*, *17*, 1820–1829.

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *PNAS*, *96*, 8301–8306.

Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behavioral and Brain Sciences*, *8*, 567–588.

Gollwitzer, P. M. & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology, 73,* 186.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews. Neuroscience*, *9*(12), 934–46.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390–2.

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

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

Haggard, P., & Tsakiris, M. (2009). The Experience of Agency. *Current Directions in Psychological Science*, *18*(4), 242–246.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage*, *50*(3), 1313–9.

Harleß, E. (1861). Der Apparat des Willens. *Zeitschrift Für Philosophie Und Philosophische Kritik*, *38*, 499 – 507.

Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *Quarterly Journal of Experimental Psychology*, *60*(11), 1540–54.

Herwig, A., & Waszak, F. (2009). Intention and attention in ideomotor learning. *Quarterly Journal of Experimental Psychology*, *62*(2), 219–27.

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology, 43,* 302-313.

Inase, M., Tokuno, H., Nambu, A., Akazawa, T., & Takada, M. (1999). Corticostriatal and corticosubthalamic input zones from the presupplementary motor area in the macaque monkey: Comparison with the input zones from the supplementary motor area. *Brain Research*, *833*, 191–201.

Jaffard, M., Longcamp, M., Velay, J.-L., Anton, J.-L., Roth, M., Nazarian, B., & Boulinguez, P. (2008). Proactive inhibitory control of movement assessed by event-related fMRI. *NeuroImage*, *42*(3), 1196–206.

Jahfari, S., Stinear, C. M., Claffey, M., Verbruggen, F., & Aron, A. R. (2010). Responding with restraint: what are the neurocognitive mechanisms? *Journal of Cognitive Neuroscience*, *22*, 1479–1492.

James, W. (1890). *The principles of psychology (Vols. 1 & 2)*. New York: Dover.

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. *Experimental Brain Research*, *192*(3), 527–32.

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. *PNAS*, *101*(36), 13335–40.

Kapur, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry*, *160*, 13–23.

Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain*, *122*, 981–991.

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., & Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *European Journal of Neuroscience*, *10*, 1209–1213.

Krieghoff, V., Brass, M., Prinz, W., & Waszak, F. (2009). Dissociating what and when of intentional actions. *Frontiers in Human Neuroscience*, *3*(February), 3.

Krieghoff, V., Waszak, F., Prinz, W., & Brass, M. (2011). Neural and behavioral correlates of intentional actions. *Neuropsychologia*, *49*(5), 767–76.

Kritikos, A., Breen, N., & Mattingley, J. B. (2005). Anarchic hand syndrome: Bimanual coordination and sensitivity to irrelevant information in unimanual reaches. *Cognitive Brain Research*, *24*, 634–647.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: how the “veto-area” exerts control. *Human Brain Mapping*, *30*(9), 2834–43.

Kühn, S., Haggard, P., & Brass, M. (2014). Differences between endogenous and exogenous emotion inhibition in the human brain. *Brain Structure & Function*, *219*(3), 1129–38.

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

Li, C. R., Huang, C., Constable, R. T., & Sinha, R. (2006). Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *The Journal of Neuroscience*, *26*, 186–192.

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*, *106 (3)*, 623–642.

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.

Lotze, R. H. (1852). *Medizinische Psychologie oder Physiologie der Seele.* Leipzig, Germany: Wedmann'sche Buchhandlung.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, *106*(1), 3–19.

Mischel, W. (1974). Processes in Delay of Gratification. *Advances in Experimental Social Psychology*, *7*, 249–292.

Mischel, W. & Baker, N. (1975). Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology, 31,* 254 - 261.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., … Shoda, Y. (2011). "Willpower" over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*(2), 252–6.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition*, *21*(1), 546–61.

Morsella, E., Berger, C. C., & Krieger, S. C. (2010). Cognitive and neural components of the phenomenology of agency. *Neurocase*, 1–22.

Morsella, E., Molapour, T., & Lynn, M. T. (2013). The three pillars of volition: phenomenal states, ideomotor processing, and the skeletal muscle system. In H. S. Terrace & J. Metcalfe (Eds.), *Joint Action and Agency*. New York: Oxford University Press.

Nachev, P., & Husain, M. (2010). Action and the fallacy of the "internal": Comment on Passingham et al. *Trends in Cognitive Sciences*, *14*(5), 192 – 193.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856–69.

Neubert, F.-X., Mars, R. B., Buch, E. R., Olivier, E., & Rushworth, M. F. S. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *PNAS, 107,* 13240–13245.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 241 - 268.

Pacherie, E. (2007). The anarchic hand syndrome and utilization behavior: A window onto agentive self-awareness. *Functional Neurology*, *22*, 211–217.

Passingham, R. E. (1993). *The frontal lobes and voluntary action. Oxford psychology series, No. 21*. New York, NY: Oxford University Press.

Passingham, R. E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14*, 16–21.

Passingham, R. E., & Wise, S. P. (2012). *The neurobiology of the prefrontal cortex: Anatomy, evolution, and the origin of insight (No. 50).* Oxford University Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.

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*, *22*(5), 613–8.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences, 8*(9), 410-417.

Sato, A. & Yasuda, A. (2003). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition, 94,* 241-255.

Schmeichel, B. J., & Zell, A. (2007). Trait self-control predicts performance on behavioral tests of self-control. *Journal of Personality*, *75*(4), 743 – 756.

Shadlen, M. N., & Gold, J. L. (2004). The neurophysiology of decision making as a window on cognition. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (pp. 1229 – 1243). MIT Press.

Sharp, D. J., Bonnelle, V., De Boissezon, X., Beckmann, C. F., James, S. G., Patel, M. C., & Mehta, M. A. (2010). Distinct frontal systems for response inhibition, attentional capture, and error processing. *PNAS*, *107*, 6106–6111.

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, *117*(11), 2341–56.

Swann, N. C., Cai, W., Conner, C. R., Pieters, T. a, Claffey, M. P., George, J. S., … Tandon, N. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *NeuroImage*, *59*(3), 2860–70.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*, 219–239.

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.

Verbruggen, F., & Aron, A. (2010). Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *PNAS*, *107*(31), 13966 – 13971.

Verbruggen, F., Chambers, C. D., & Logan, G. D. (2013). Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. *Psychological Science*, *24*(3), 352–62.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418–424.

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neurosci Biobehav Rev*, *33*(5), 647–661.

Vohs, K. D., & Schooler, J. W. (2008). The Value of Believing in Free. *Psychological Science*, *19*(1), 49–54.

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), 346–56.

Webb, T. L. & Sheeran, P. (2003). Can implementation intentions help to overcome ego-depletion? *Journal of Experimental Psychology, 39,* 279 - 286.

Wessel, J. R., & Aron, A. R. (2014). Inhibitory motor control based on complex stopping goals relies on the same brain network as simple stopping. *NeuroImage*, *103*, 225–234.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

Voluntary inhibition of pain avoidance behavior: An fMRI study[1]

Behavioral inhibition has classically been considered to rely upon a neural network centered at the right inferior frontal cortex (rIFC; Aron, Robbins, & Poldrack, 2004; 2014). However, the vast majority of inhibition studies have entailed exogenous stop signals instructing participants to withhold responding. More recent work has begun to examine the neural underpinnings of endogenous inhibition, revealing a distinct cortical basis in the dorsal fronto-median cortex (dFMC; Brass & Haggard, 2007; Kühn, Haggard, & Brass, 2009). Yet, contrary to everyday experiences of voluntary behavioral suppression, the paradigms employed to investigate action inhibition have thus far been somewhat artificial, and involve little persuasive motivation to act. Accordingly, the present fMRI study seeks to compare and contrast intentional with instructed inhibition in a novel pain paradigm that recruits 'hot' incentive response systems. Participants received increasing thermal stimulation to their inner wrists, and were required to occasionally withhold their natural impulse to withdraw from the compelling pain sensation at peak temperature, in both instructed and free-choice conditions. Consistent with previous research, we observed inhibition-related activity in the dFMC and the rIFC. However, these regions displayed equivalent activation levels for both inhibition types. These data extend previous research by demonstrating that under ecologically valid conditions with a strong motivation to act, both stopping networks operate in concert to enable suppression of unwanted behavior.

[1] In collaboration with Jelle Demanet, Ruth M. Krebs, Pieter Van Dessel, & Marcel Brass (under revision, Brain Structure and Function).

INTRODUCTION

Self-control permits individuals to fulfill their goals, keep their promises, and conform to societal norms (Baumeister, Vohs, & Tice, 2007). Successful self-regulation often depends on the ability to suppress the urge to act impulsively (Muraven & Baumeister, 2000). This can be witnessed in everyday examples ranging from the relatively innocuous (e.g., abstaining from eating a second donut) to the potentially dire (e.g., refraining from striking another person in a fit of rage). While failure at the former example might prove disappointing for the individual, poor impulse control as displayed in behaviors stemming from violence, addiction, or unconstrained emotion can be particularly damaging on a broad social scale. The ability to successfully delay gratification and inhibit one's urges is predictive of long-term outcomes such as job success, health, self-esteem, and helping behavior/aggression (Mischel et al., 2011; Tangney, Baumeister, & Boone, 2004). Crucially, this control must originate within the person for it to be effective repeatedly and over time; merely adhering to external rules and laws is insufficient when high levels of regulatory effort must be exerted.

At the neural level, behavioral inhibition has classically been considered to rely upon a network centered at the right inferior frontal cortex (rIFC; see Aron, Robbins, & Poldrack, 2004; 2014 for reviews). In the vast majority of these studies, participants were instructed to withhold a response upon the appearance of an external stop signal (e.g., Aron & Poldrack, 2006; Chikazoe et al., 2007; 2009; Garavan et al., 1999; 2002). Yet, in naturalistic settings, inhibition of unwanted behavior is typically based on internal decisions rather than on external instructions. For instance, a person who is trying to give up smoking will not rely on an external stop signal, but rather on their internal willpower in order to resist the temptation to light a cigarette (Muraven, 2010). More recent studies have begun to address this issue by adding a choice condition in which participants freely choose between performing or withholding a response (see Filevich et al., 2012 for a review). These studies have revealed a distinct cortical basis for endogenous[2] inhibition in the dorsal fronto-median cortex (dFMC; Brass & Haggard, 2007; 2008; Kühn, Haggard, & Brass, 2009). In their 'What-When-Whether' model of intentional action, Brass & Haggard (2008) proposed that the dFMC may constitute a

[2] We use the term 'endogenous' synonymously with 'voluntary' and 'intentional' to denote an internal locus of the decision to perform or withhold an action.

'veto area' serving a final decision as to whether or not to execute an already prepared action plan, making this brain region essential for the exertion of self-control. Converging evidence suggests that these "veto" signals might be implemented by the preSMA, which has been shown to be functionally connected with the dFMC during intentional inhibition (Kühn et al., 2009; 2013) and also plays a pivotal role in externally-guided response inhibition, presumably via direct and indirect projections to the motor cortex (Cai et al., 2012; Nachev, Kennard, & Husain, 2012; Swann et al., 2012).

However, although these studies were able to extend the scope of prior studies involving purely externally-guided response inhibition, the paradigms have thus far employed relatively artificial experimental settings, lending participants little prior motivation to act or inhibit. Decisions were thus based on rather arbitrary choices and arguably might not have required genuine self-control. It therefore remains an open question whether the distinction between externally triggered and intentional inhibition is also valid in situations where a very strong response tendency has to be inhibited. One can argue that in such situations strong intentional control is required regardless of whether one decides to inhibit or is externally cued. In the present study, we sought to address this question and examine the neural basis of intentional inhibition in a more ecologically-valid setting that recruits 'hot' incentive response systems (Metcalfe & Mischel, 1999). Pain was selected as the behaviorally relevant stimulus for our purposes, as the organism is strongly motivated to avoid the pain sensation (Campbell & Misanin, 1969; Elliot, 2006). Moreover, despite its strong prepotency, the pain avoidance response can at times be voluntarily suppressed when higher-order goals call for such behavior (cf. Morsella, 2005). Accordingly, management of the pain avoidance response can be seen as a classical instance of self-control over one's basic drives. In the present study, participants received thermal stimulation to alternating inner wrists, and were required to occasionally withhold their natural impulse to withdraw from the compelling pain sensation at peak temperature, in both instructed and free-choice conditions. This allowed us to investigate the neural basis of endogenous and exogenous inhibition of behavior under the presence of a strong urge to act. Based on the aforementioned findings, we expected the dFMC to be involved in endogenous inhibition, and the rIFC in exogenous inhibition. In addition, we wanted to explore whether or not the previously reported independence of the neural networks

involved in these two forms of inhibition still holds under motivationally salient conditions.

METHODS AND MATERIALS

Participants

Twenty-one native Dutch speakers (7 males) participated in the study (mean age = 22.2 years, SD = 3.6); each reported as healthy and had no history of neurological, pain, or circulatory disorders. The data from the second run of participant 6 were excluded prior to analysis due to excessive head movements (> 5 mm). All participants gave written informed consent, and the study was approved by the Medical Ethical Review Board of the Ghent University hospital, in accordance with the declaration of Helsinki. Participants were right-handed, as assessed by the Edinburgh Inventory (Oldfield, 1971), and were compensated thirty-five euros for their participation.

Experimental Procedure

Pain tolerance threshold determination. Pain was induced via a thermode connected to a Medoc PATHWAY device (MEDOC, Haifa, Israel), an apparatus designed for the induction of thermal pain using cold or hot stimulation. The temperature at which participants felt a sufficient amount of pain was determined during a pre-test session taking place one week prior to scanning. Participants were exposed to 26 trials in which the thermal sensation gradually increased over five seconds from 32°C to a randomized destination temperature between 45 and 50°C (in increments of .25 degrees), a slope comparable to the experimental trials. After each trial, the thermode returned instantly to baseline temperature, and participants were asked to rate their perceived pain on a scale from zero to eight, with zero being no pain and eight being the worst possible pain. The destination temperature employed in the main experiment was computed for each participant as the highest temperature at which they rated their pain as a six[3]. This method was revealed during piloting to yield more accurate tolerance threshold measurements than merely requiring participants to indicate the maximum heat they could withstand when exposed to a steadily increasing

[3] This temperature is henceforth referred to as the individual participant's tolerance threshold.

temperature. Importantly, participants were free to press a button at any point during the threshold determination in order to terminate the trial.

***Task and stimuli*.** Participants received thermal pain stimulation during each trial, applied via a thermode to alternating inner wrists. The images of three geometric shapes (triangle, square, circle) were used as cues to indicate the trial type. Depending on the cue, participants were requested to select one of the following response options: press the button as quickly as possible in order to terminate the trial ('directed action,' 25% of trials), inhibit this response and endure the pain for an additional two seconds ('directed inhibition,' 25% of trials), or make a voluntary decision to either button press immediately or persist ('choice action' and 'choice inhibition,' respectively, combined equaling 50% of trials). In the latter case, participants were requested to choose both options approximately equally often over the course of the experiment, but not to use any particular strategies (e.g., simple response alternations), or to decide in advance of the presentation of the cue. Adherence to these instructions was subsequently assessed by calculating each participant's Random Number Generation 2 (RNG2; an index optimized for two-choice response sequences) score using the program RgCalc (Towse & Neil, 1998). RNG2 scores range from 0 to 1, with 1 indicating complete sequence predictability. A pilot study had revealed that participants are typically around 200 ms slower to respond on choice action trials than on directed action trials, presumably reflecting the additional time needed for the decision process. Accordingly, to make stimulation as similar as possible across action conditions, 200 ms of thermal stimulation was added to directed action trials, following the button press.

Each trial began with the presentation of a fixation cross for five seconds, during which time the temperature of the thermode began to gradually increase from a baseline of 32°C to the participant's individually determined tolerance threshold. Subsequently, one of the three task cues appeared in place of the fixation cross, and persisted for the remainder of the pain stimulation. The temperature remained at tolerance threshold for the next two seconds, or until the participant pressed the button to terminate both the pain stimulation and the trial. Participants responded with the index or middle finger of the arm not being stimulated (thereby providing a response time for action trials). This was followed by a six second rest period. Afterwards, prompts were presented for two seconds each, asking participants for

verbal ratings (collected via a microphone inside the scanner bore) of their perceived pain and their subjective 'urge to terminate the trial by pressing the button' (both on a scale of 0 – 8). Participants were then cued to alternate the arm placed atop the thermode, and were given 10 seconds in which to accomplish this task with the assistance of an experimenter who stood near the scanner bore. The experimenter also placed a small sandbag over the to-be stimulated wrist in order to lend weight and prevent the participant from inadvertently withdrawing from the pain source rather than button pressing. Each trial ended with an additional six-second rest period. A schematic overview of a possible trial in the choice condition is presented in Figure 1. The assignment of geometric shapes to trial types was counterbalanced across subjects. Each participant had to perform eighty trials in total, being divided into two runs of forty trials presented in randomized sequence. In each run, participants were given twenty trials in which they were cued to make a decision, ten trials in which they were cued to push and ten trials in which they were cued to inhibit their withdrawal response. Importantly, participants were free to press a button to immediately terminate the thermal sensation at any point during the experiment.

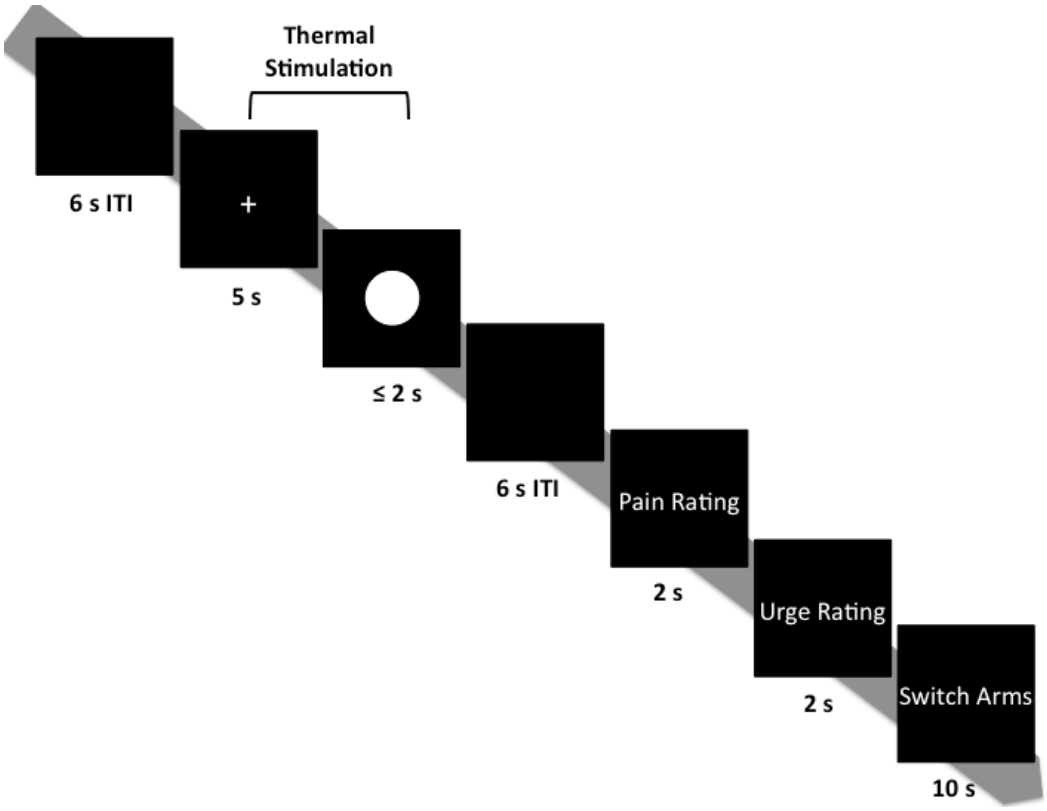

Figure 1. ***Schematic of a sample trial in the choice condition***

Behavioral data preparation

Errors in the form of responding on a directed inhibition trial led to the exclusion of 0.77% of trials (13 out of 1680 total trials; number of excluded trials per participant: *M* = 0.62, *SD* = 0.92) from both behavioral and fMRI analysis. No errors were committed via failing to press the response button on a directed press trial. Pain and urge ratings were analyzed using a repeated-measures ANOVA with INSTRUCTION (directed vs. choice) and RESPONSE (action vs. inhibition) as within-subjects factors. Reaction times on action trials were analyzed as a function of INSTRUCTION via two-tailed paired-samples *t*-tests.

fMRI data acquisition and preprocessing

Data were acquired with a 3T Siemens Magnetom Trio MRI system (Siemens Medical Systems, Erlangen, Germany) using a 32-channel radiofrequency head coil. Subjects were positioned headfirst and supine in the magnet bore. First, 176 high-resolution anatomical images were acquired using a T1-weighted 3D MPRAGE sequence (TR = 2,250 ms, TE = 4.18 ms, TI = 900 ms, image matrix = 256 x 256, FOV = 256 mm, flip angle = 9°, and voxel size = 1 x 1 x 1 mm). Whole-brain functional images were then collected using a T2-weighted echo-planar imaging (EPI) sequence, sensitive to blood-oxygen-level dependent contrast (TR = 2,000 ms, TE = 35 ms, image matrix = 64 x 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 17%, voxel size 3.5 x 3.5 x 3.0 mm, and 30 axial slices). A varying number of images were acquired per run due to individual differences in choice behavior and reaction times.

All data were preprocessed and analyzed using Matlab and the SPM8 software (Wellcome Department of Cognitive Neurology, London, UK). To account for possible T1 relaxation effects, the first four scans of each EPI series were excluded from the analysis. First, a mean image for all scan volumes was created, to which individual volumes were spatially realigned using rigid body transformation. Thereafter, they were slice time corrected using the first slice as a reference. The structural image of each subject was coregistered with their mean functional image after which all functional images were normalized to the Montreal Neurological Institute (Montreal, Quebec, Canada) T1 template. The images were resampled into 3.5 x 3.5 x 3.5 mm voxels and spatially smoothed with a Gaussian kernel of 8 mm (full-width at half maximum). A

high-pass filter of 128 Hz was applied during fMRI data analysis.

Statistical analyses

The first-level statistical analyses were performed using a general linear model (GLM, Friston et al., 1995). Of primary interest were the brain regions involved in the implementation of intentional inhibition, or the decision to intentionally inhibit. We therefore used the onset of the task cue as the main event of interest in the GLM. It is important to note that all trials were identical in terms of stimulation up to and including cue onset, and for an average of 750 ms afterwards (i.e., average time it took to respond in action trials). Based on the factorial design, four regressors were defined reflecting the experimental conditions ('directed action,' 'directed inhibition,' 'choice action,' and 'choice inhibition'). Temporal derivatives of these regressors were added to the model, and six additional regressors defining head movements were also included to account for any residual movement-related effects. All regressors were convolved with a canonical hemodynamic response function (HRF). Contrast images were computed separately for each participant to compare parameter estimates of the relevant conditions.

These contrast images were advanced to the second level, using a random-effects within-subject flexible-factorial design as implemented in SPM8 with factors INSTRUCTION (directed vs. choice) and RESPONSE (action vs. inhibition). First, we examined the main effects to reveal brain areas that are (i) more active on inhibition trials than on action trials, and (ii) more active on choice trials than on directed trials. In a second step, we directly contrasted intentional inhibition with externally guided inhibition to reveal the brain areas that are specifically engaged in the internal decision to inhibit the pain avoidance response. To control for false-positive rates, combined voxel activation intensity and cluster extent thresholds corrected for multiple comparisons were determined using 3dClustSim (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html). This widely used correction method is applied to statistical contrast images at the group level and estimates the probability of observing false positive (random fields of noise) clusters of a given size, as a function of a given voxelwise *p* value. The 3dClustSim program considers the size of the image (number of voxels), the voxelwise statistical values, and

the spatial correlations over voxels (spatial smoothness) and runs a user-specified number of Monte Carlo simulations to generate an appropriate null-distribution. Here, ten thousand Monte Carlo simulations were run, taking into account the whole-brain search volume and the estimated smoothness of each axis (*x, y,* and *z*) of the respective group SPMs. Probability estimates of a random field of noise were generated, producing a cluster of voxels of a given extent for a set of voxels passing a voxelwise *p* value threshold of 0.001. Given this voxelwise threshold, the simulations determined that cluster sizes of 20.0 – 23.5 voxels, depending on the specific contrast analysis, corresponded to a combined threshold of $p < 0.05$ (corrected). Whole-brain analyses were supplemented with region-of-interest (ROI) analyses. ROIs were generated using MARSBAR toolbox for use with SPM 8 (Brett, Anton, Valabregue, & Poline, 2002).

RESULTS

Behavioral results

Proportion of choices. For choice trials, participants were able to follow the instruction to choose both options equally often, and inhibited their response on average in 49.9% of the trials (Range = 37.5% - 65%; *SD* = 7.3%).

Randomness of choice response sequences. Participants displayed a mean RNG2 index of 0.724 (*SD* = .008). Individual scores were compared to twenty-one randomly generated sets of two-choice response sequences (RNG2 *M* = 0.723, *SD* = .007) in an independent samples *t*-test. Participants' choice trial responses did not differ significantly from the randomly generated samples [$t(40) = .376$, $p = .709$], suggesting that they did not use simple alternation strategies as a means of conforming to the experimental instructions.

Subjective pain ratings. The analysis of pain ratings revealed main effects of INSTRUCTION [$F(1,20) = 7.205$, $p = .014$, $\eta p = .265$] and RESPONSE [$F(1,20) = 43.479$, $p < .001$, $\eta p = .685$] as well as a significant interaction [$F(1,20) = 5.591$, $p = .028$, $\eta p = .218$]. Participants rated inhibit trials as more painful than action trials (presumably due to the greater stimulation duration), with directed inhibit trials receiving the highest pain ratings (Directed Action: *M* = 4.24, *SE* = 0.25; Choice Action: *M* = 4.28, *SE* = 0.25; Directed Inhibition: *M* = 5.50, *SE* = 0.21; Choice Inhibition: *M* = 5.03, *SE* = 0.24). A post-

hoc *t*-test revealed that pain ratings do not differ between directed action and choice action trials, $t(20) = .217$, $p = .831$.

Subjective urge ratings. The analysis of urge ratings revealed a significant interaction of INSTRUCTION and RESPONSE [$F(1,20) = 7.194$, $p = .014$, $\eta p = .265$], driven by the choice condition and reflecting that participants experienced similar urges on directed trials, but when given a choice, their urges were concordant with their decisions (i.e., a higher urge to press on action trials). No main effects were significant (Directed Action: $M = 4.00$, $SE = 0.34$; Choice Action: $M = 4.23$, $SE = 0.34$; Directed Inhibition: $M = 3.78$, $SE = 0.33$; Choice Inhibition: $M = 3.24$, $SE = 0.24$).

Objective thermal stimulation. Due to the constraints of the Medoc software, some slight variation in thermode temperature is inherent throughout the course of the experiment (see Figure 2), yet participants did not appear to use either temperature at cue onset or peak temperature (typically occurring 100 ms after cue onset) as factors in their decision to press or inhibit. Mean thermode temperatures were computed per participant at both cue onset and thermal peak for each of the choice conditions. Two-tailed paired-samples *t*-tests revealed that mean temperature at onset and peak did not differ significantly between choice inhibit and press trials, $t(20) = .037$, $p = .971$ and $t(20) = .142$, $p = .889$ respectively. Grand mean onset temperature was 49.23 degrees Celsius ($SD = 0.45$), while grand mean peak temperature was 50.00 degrees Celsius ($SD = 0.47$). Participants' tolerance thresholds did not correlate with reaction times, mean reported pain, or proportion of inhibition on choice trials (all ps > .169).

Reaction times. As expected, response times were significantly slower on choice trials than on directed trials, $t(20) = 6.68$, $p < .001$ (directed = 597 ms, choice = 903 ms), suggesting that participants did not make their decisions in advance of the cue.

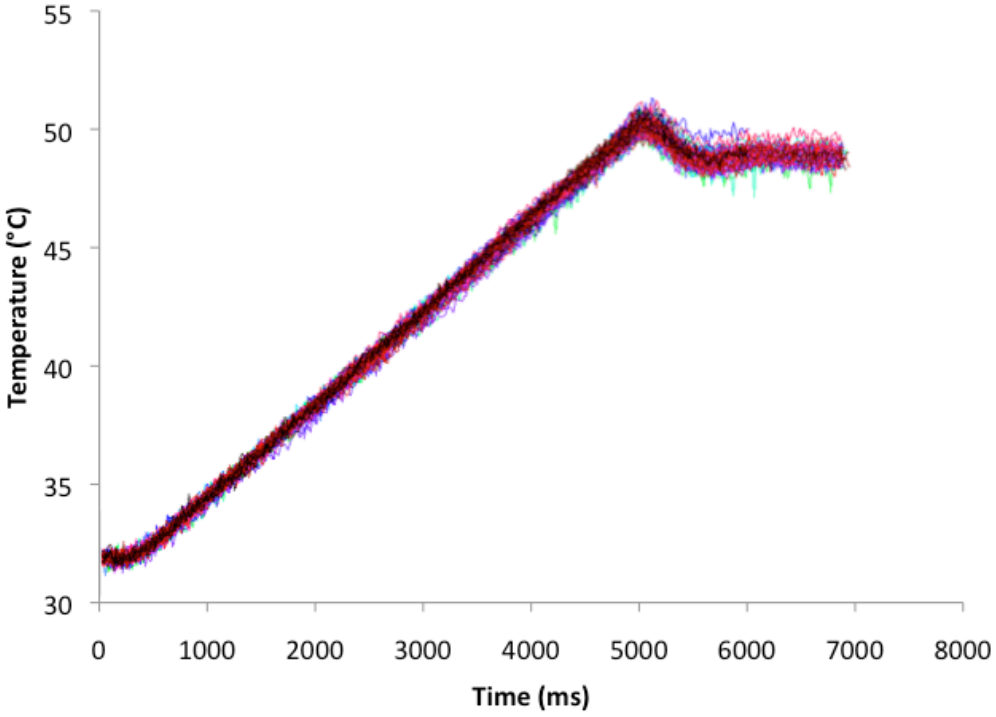

Figure 2. ***Composite of all 80 trials for an exemplar participant, representing actual stimulation temperatures. Note each trial is displayed in a distinct color and the overlap of the slope indicates equivalent stimulation across experimental trials.***

fMRI results

Whole-brain analyses. We first analyzed the main effect of RESPONSE to reveal the brain areas that exhibit stronger activation on inhibition trials than on action trials (see Figure 3). As expected, activity was found within the right inferior frontal gyrus (rIFG) extending into the adjacent insular cortex. The location was similar to that observed in previous studies investigating externally guided inhibition (e.g., Aron & Poldrack, 2006; MNI x, y, z = 44 12 8). Another cluster was located in the anterior medial PFC (amPFC), encompassing the dFMC coordinates reported in previous studies on intentional inhibition (Brass & Haggard, 2007; Kühn et al., 2009). Two additional clusters were located in the ventral part of the anterior supplementary motor area (pre-SMA) and the right inferior parietal lobule (IPL), both areas that have previously been implicated in intentional and externally-guided action inhibition (e.g., Filevich et al., 2012; Kühn et al., 2009). Secondly, we analyzed the main effect of INSTRUCTION in

order to identify brain regions that show increased activity on choice trials compared with directed trials (Figure 3). Activity was found in the medial frontal wall, extending from the pre-SMA and with the majority of activation located within the rostral cingulate zone (RCZ), bilaterally in the dorsolateral prefrontal cortex (dlPFC), and in the right IPL. This is in accordance with previous descriptions of a frontoparietal 'choice network' engaged in internal action decisions (e.g., Haggard, 2008; Mueller et al., 2007). Finally, we contrasted 'choice inhibition' trials with 'directed inhibition' trials to examine which brain areas are specifically involved in intentional (vs. externally-guided) inhibition. This analysis revealed a single cluster of activity in the right dorsal pre-SMA. The reverse contrast (directed versus choice inhibition) yielded no significant activity. For a complete overview of significant activations, see Table 1.

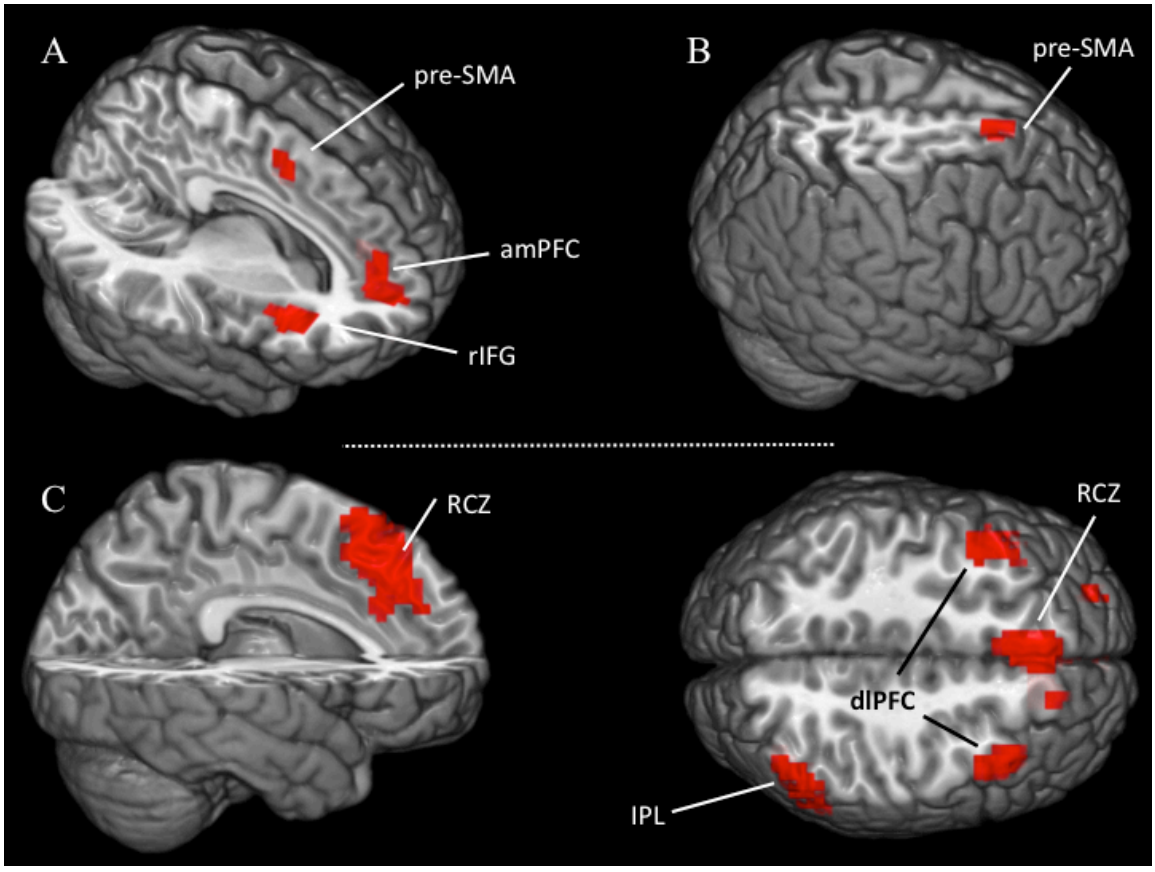

Figure 3. ***Activation maps of the whole-brain contrasts, comparing a) inhibition versus action; b) choice inhibition versus instructed inhibition; and c) choice versus instructed trials. Note activation maps were thresholded at p < .001 (uncorrected), with a minimum cluster size of 22 contiguous voxels (see Methods section for details).***

Table 1

Overview of activation clusters revealed by the whole-brain analyses. Montreal Neurological Institute (MNI) coordinates reflect the peak voxel of a given cluster.

Region	Voxels	MNI coordinates			*Tmax*
		x	*y*	*z*	
Main effect inhibition: inhibit > press					
Right inferior frontal gyrus	127	41	24	6	5.26
Anterior insula		38	10	10	4.33
Anterior median prefrontal cortex	77	10	46	10	4.50
Inferior parietal lobule	26	52	-32	30	4.08
Pre-supplementary motor area	28	10	4	48	4.00
Main effect choice: choice > directed					
Pre-supplementary motor area	500	16	18	66	5.82
Rostral cingulate zone		-4	38	34	5.10
Rostral cingulate zone		6	32	30	5.09
Left dlPFC	77	-43	10	48	5.53
Left inferior frontal gyrus	67	-50	24	-1	4.69
Superior frontal gyrus	23	-22	60	24	3.69
Right dlPFC	57	44	21	48	5.05
Inferior parietal lobule	122	55	-56	38	4.94
Choice inhibit > directed inhibit					
Pre-supplementary motor area	31	16	18	66	4.66
Effect of pain level, median split: high pain > low pain					
Anterior cingulate cortex	41	-1	10	41	3.85
Precentral gyrus	28	-60	-4	10	4.54
Main effect action: press > inhibit					
Right motor cortex	352	48	-21	62	11.22
Left motor cortex	302	-46	-21	62	7.55

Region-of-interest (ROI) analyses. Based on the results of the whole-brain analysis, we conducted an additional series of ROI analyses. In the first set, we were interested in further examining the role of the dFMC. Given the absence of a difference in dFMC activity between the two inhibition conditions in the whole-brain analysis, we sought to test whether there was a sub-threshold difference between inhibition conditions that could not be detected in the whole-brain contrast. To this end, we generated a spherical ROI (radius = 10 mm, MNI peak -7 42 21) centered on the peak dFMC coordinate reported by Kühn et al. (2009) in order to provide an independent non-circular analysis (see Kriegeskorte et al., 2009). This analysis yielded a significant main effect of INSTRUCTION, $F(1,20) = 9.663$, $p = .006$, reflecting increased activity on choice trials compared with directed trials. The effect of RESPONSE was significant as well, $F(1,20) = 4.604$, $p = .044$, indicating greater activation on inhibition trials than on action trials[4]. Moreover, there was a significant interaction between the factors, $F(1,20) = 4.554$, $p = .045$ (see Figure 4). Post-hoc comparisons revealed, however, that the interaction was driven by increased activity on choice action trials compared with directed action trials, $t(20) = 3.689$, $p = .001$, whereas activity did not differ between the two inhibition conditions, $t(20) = 1.000$, $p = .329$. Altogether, the above ROI analysis confirms that in the context of pain avoidance behavior, the dFMC showed reliable activation related to behavioral inhibition, but did not furthermore depend upon the level of instruction.

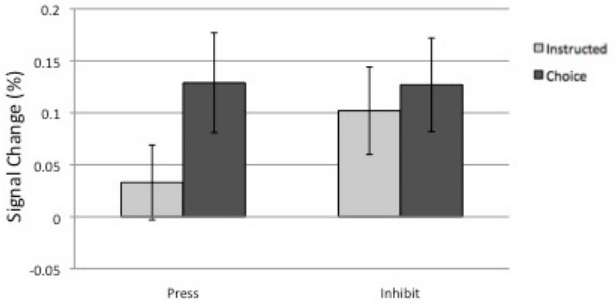

Figure 4. ***Results of the frontomedian ROI analysis. A spherical ROI (radius = 10mm) was centered at the peak dFMC coordinate (-7 42 21) derived from the study of Kühn et al. (2009). Values represent the mean percent signal change (beta values) and standard errors.***

[4] A similar pattern of results was observed with a spherical ROI centered on the peak amPFC coordinate derived from the present study. However, these results are not displayed here due to the non-independence of ROI selection and analysis.

The second set of ROI analyses elaborated on activity in the right insula, an area that has been implicated in pain processing (e.g., Ingvar, 1999). In particular, we sought to examine to what extent inhibition-related activity in this region was in fact driven by the demand to inhibit a response rather than merely a consequence of the lengthier thermal stimulation on these trials, compared with press trials. With this in mind, we performed two analyses. In a first step, we defined an ROI corresponding to the right insula based on anatomical criteria drawn from the Anatomy toolbox for SPM (Eickhoff et al., 2005). Thereafter, we computed each participant's mean pain ratings separately for each experimental condition and correlated these ratings with the respective percent signal change within the ROI. Correlations were Bonferroni-corrected for multiple comparisons. If activity in the insula was driven by differences in the amount of pain, then it should co-vary with the subjective ratings regardless of the experimental condition. Yet no such correlations were found between mean pain ratings and percent signal change within any of the four conditions, *p*s > .240, suggesting that activity in the insula was driven by the demand to inhibit a pain avoidance response and not by mere differences in pain levels.

To further elaborate on this conclusion, we performed an additional whole-brain analysis contrasting high-pain trials with low-pain trials across all experimental conditions, based on a median split of the individual pain ratings. This analysis should reveal brain areas that are generally sensitive to the level of subjective pain, irrespective of the experimental condition. Indeed, the anterior cingulate cortex (ACC, MNI peak at -1 10 41) and precentral gyrus (MNI peak at -60 -4 10) were the only regions to exhibit stronger activity on high- versus low-pain trials; the ACC has frequently been implicated in processing of pain (Rainville, Duncan, Price, Carrier, & Bushnell, 1997; Wager et al., 2004). In contrast, no difference was observed in the right insula, confirming our conclusion that activity of this region was driven by the demand to withhold a response and not by different levels of pain (see Figure 5).

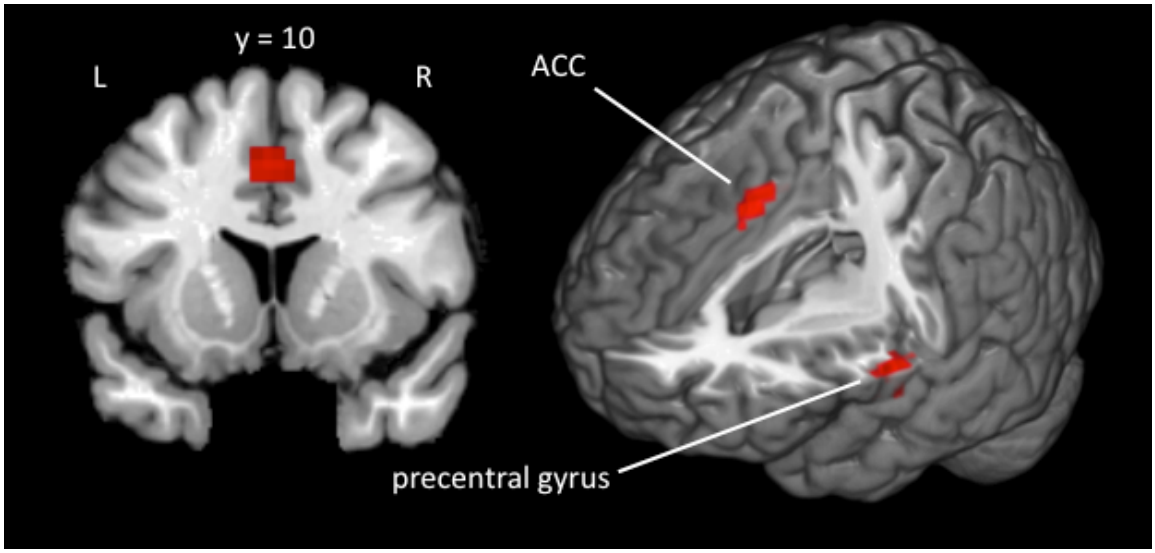

Figure 5. ***Main effect of high versus low pain, across conditions. Note activation maps were thresholded at p < .001 (uncorrected), with a minimum cluster size of 22 continguous voxels (see Methods section for details).***

We conducted a third set of ROI analyses to address the post-hoc question of how inhibition in our paradigm should be characterized; does inhibition target the response finger that can terminate the thermal stimulation via button press or instead the stimulated hand (thereby maintaining the posture of the stimulated wrist and avoiding withdrawal)? To answer this question, we first defined bilateral ROIs of the motor hand area based on the main effect of action reported in Table 1. Spherical ROIs with a radius of 10mm were defined over the peak left and right motor cortex coordinates (left: -46 -21 62; right: 48 -21 62). Beta values were collapsed across hemispheres (i.e., a right hand response was coded as the stimulated hand and a left hand response as the effector hand in the right hemisphere, while a left hand response was coded as the stimulated hand and a right hand response as the effector hand in the left hemisphere) and subjected to a repeated-measures ANOVA with with factors RESPONSE (Action vs. Inhibition) and HAND (Effector vs. Stimulated). This analysis revealed a significant main effect of RESPONSE, $F(1, 20) = 107.987$, $p < .001$, reflecting greater activity on action trials, a non-significant main effect of HAND, $F(1, 20) = .335$, $p = .569$, and a significant interaction, $F(1, 20) = 131.724$, $p < .001$ (see Figure 6). We conducted post-hoc *t*-tests between the effector hand and the stimulated hand on both

trial types to specify the interaction. As expected, on action trials activity was greater in the effector hand than in the stimulated hand, $t(20) = 8.071$, $p < .001$. Importantly, however, on inhibition trials activity in the effector hand was reduced when compared with the stimulated hand, $t(20) = 5.718$, $p < .001$. Overall, activity did not differ between action trials and inhibition trials in the stimulated hand, $t(20) = .219$, $p = .829$. These results suggest that inhibition in our paradigm was directed only towards the instrumental response that could terminate the pain indirectly via a button press, rather than towards the "hard-wired" withdrawal response of the stimulated hand. It should be further noted that the weight of the sandbag placed over the stimulated wrist would make the latter response option quite difficult.

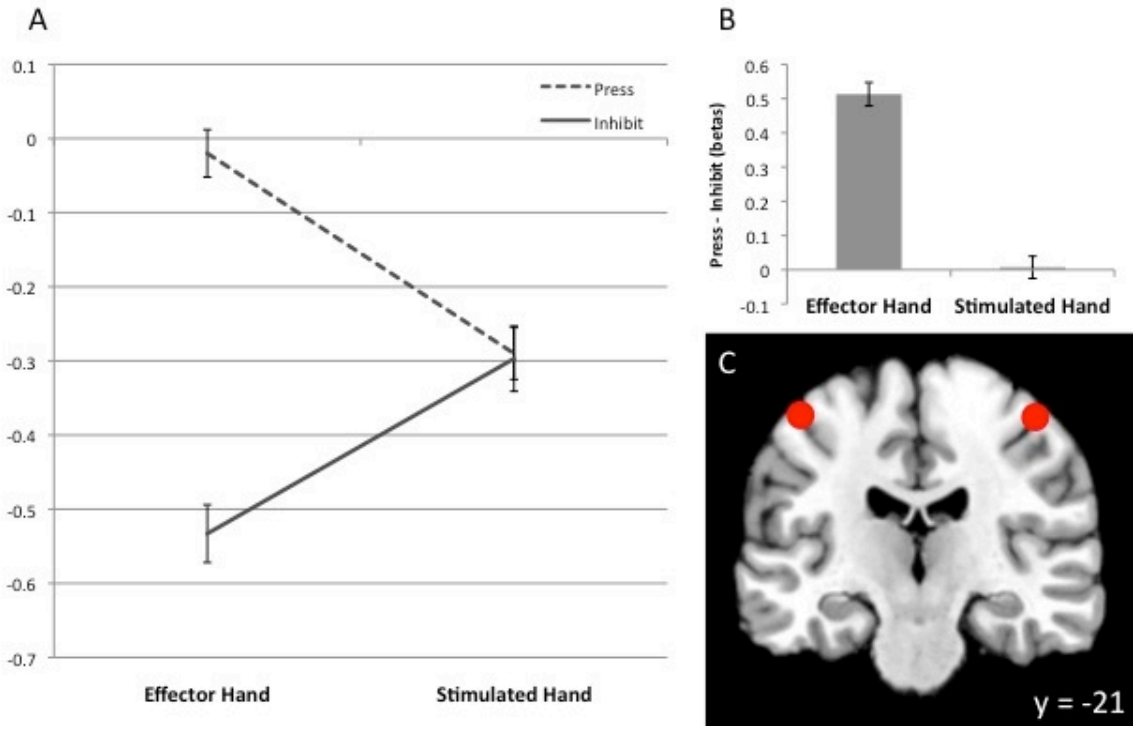

Figure 6. ***A) Motor cortex activation pattern for effector and stimulated hands. Values represent the mean percent signal change (beta values) and standard errors. B) Activation difference scores (press – inhibit) for each hand. C) Locations of the peak coordinates used to draw motor cortex ROIs.***

DISCUSSION

The present study sought to investigate the neural basis of intentional inhibition in the context of pain avoidance behavior. Using thermal stimulation permitted us to

create an experimental setting that requires genuine self-control in order to withhold responding, representing a major step towards a more ecologically-valid investigation of the intentional inhibition of action. In this context, we observed brain activity mirroring previous findings concerning both intentional and externally guided inhibition. There was little overt difference between inhibition types, indicating that the two forms of inhibition are more strongly linked under motivationally salient conditions. Furthermore, inhibition seemed to be specifically directed towards the effector implementing the instrumental response. Below we discuss the implications of our findings for theories on the neural basis of action inhibition and self-control.

The role of the dFMC in self-control

Our finding that the dFMC was activated more strongly for response omissions than for response executions replicates a growing body of research that has indicated the prominent role of this area in the withholding of actions (Filevich et al., 2012). As our paradigm employed quite a different procedure than previous intentional inhibition studies[5], this finding strengthens the conclusion that the dFMC is in fact critical for the voluntary cancellation of an action, independent of a particular experimental context. Yet in contrast with previous studies, we found that the dFMC showed equivalent activity for both intentional and externally-guided inhibition. Given that the urge to withdraw from a pain source is intrinsically very high, it seems plausible that the level of instruction had a reduced impact upon the recruitment of self-regulatory processes in our paradigm. Rather, suppression of the pain avoidance urge might generally require a high degree of self-control, necessitating strong intentional engagement even under externally-guided instruction.

Interestingly, the dFMC activation in our study was located slightly more rostrally than in previous inhibition studies, and was encompassed by parts of the anterior medial PFC that have been implicated in self-referential thought processes (e.g., Johnson et al., 2006; Mitchell et al., 2005). This could imply that the role of the dFMC in intentional inhibition is more general than previously assumed. We have

[5] Previous studies employed either a modified version of the 'Libet task' (Brass & Haggard, 2007) in which self-paced button presses are occasionally omitted, or a 'ramp task' (Kühn et al., 2009), also requiring participants to occasionally withhold from responding.

recently proposed that its contribution should be characterized as general disengagement from otherwise prepotent intentions, plans, and urges (Lynn, Muhle-Karbe, & Brass, 2014) rather than as pure motor suppression. Cancelling an action that has already been prepared is one such instance, requiring disengagement from the motor intention (e.g., Brass & Haggard, 2007; Kühn et al., 2009). However, similar activity has also been found when participants need to distance themselves from negative emotions (Kühn et al., 2011) or cigarette cravings (Brody et al., 2007), and to quit loss-chasing (Campbell-Meiklejohn et al. 2008). In addition, the dFMC has also been implicated in mentalizing, which is assumed to rely upon disengagement from the otherwise dominant self-perspective (e.g., Brass, Derfuss, & von Cramon, 2005; Brass, Ruby, & Spengler, 2009). We therefore believe that disengagement provides a useful integrative concept for explaining the dFMC's involvement in these diverse functions. However, as this view is thus far based primarily on reverse inference, further research will be necessary to test this idea and to reveal the precise contribution of the dFMC to self-control.

Pre-SMA

The only region that exhibited stronger activity for intentional inhibition than for externally-guided inhibition was the pre-SMA. This region has been shown to be connected with the dFMC anatomically (Johansen-Berg et al., 2004) and functionally (Kühn et al., 2009), typically with the assumption that the dFMC feeds top-down stopping signals to the pre-SMA, which in turn implements inhibition via connections with motor areas (Cai, George, Verbruggen, Chambers, & Aron, 2012; Kühn et al., 2009). The differential activation levels in the pre-SMA for the two inhibition types are therefore likely related to differences in the way that suppression of the pain avoidance response is implemented in the respective trials. The immediacy of action decisions could be of particular importance in this respect. In externally guided trials, the cue directly informs participants about the outcome, thereby triggering an instantaneous cancellation of the response (Verbruggen & Logan, 2008). In choice trials, on the other hand, participants may implement a continual delay of the response rather than a singular decision not to act, as the veto can theoretically be implemented at any time throughout the trial. Accordingly, increased activation of the pre-SMA with intentional

inhibition could be related to decision uncertainty on choice trials, and the continuous rather than instantaneous suppression of the pain avoidance response. Examining the preSMA's connectivity with other regions implicated in inhibition may help to clarify its contributions to different types of behavioral inhibition (similar to Herz et al., 2014).

Insular cortex

An interesting aspect of our results involves the activation of the right anterior insula, which we found in conjunction with the rIFG related to both types of behavioral inhibition. Given the numerous processes to which this region has been linked, a conclusive interpretation of its role in our paradigm seems difficult to obtain. Previous studies have shown that the insula might be involved in inhibitory control (e.g., Forstmann et al., 2008; Hodgson et al., 2007), yet is also a well-known component of the pain matrix (Garcia-Larrea & Peyron, 2013; Ingvar, 1999; Peyron, Laurent, & García-Larrea, 2000), and as such, activation of this region might be considered a mere consequence of the longer stimulation duration in inhibition trials compared with press trials. However, several points related to the design and analysis of our study suggest that this is not entirely the case. First, our model was based on brain activity time-locked to the instructional cue, thus a time point at which all conditions were equivalent in terms of pain stimulation, and remained so for an additional 750 ms, on average. Second, contrasting trials with high vs. low levels of subjective pain yielded only activity within the ACC, another well established component of the pain matrix (e.g., Garcia-Larrea & Peyron, 2013), but no differential activity in the insula. Third, the ROI analysis of the right insula (applying a median split based on the subjective pain ratings) revealed no significant correlations between mean pain ratings per condition and percent signal change within each region of the insula. Accordingly, we are confident that the inhibition-related activation of the insula in our study is not exclusively related to the differential levels of pain that accompanied those trials. Importantly, previous intentional inhibition studies that did not involve somatic stimulation have also reported activity in the insula (Brass & Haggard, 2007; Campbell-Meiklejohn et al., 2008; Kühn et al., 2009). Although only speculative at this point, this could imply that the role of the insula in the context of self-control reflects affective evaluation of the behavioral outcome (see Brass & Haggard, 2010).

A hierarchical model of motivated self-control

On a larger scale, the findings of the present study indicate that previous views on the structure of behavioral inhibition must be reconsidered in light of motivational factors. In particular, the co-activation of both previously described inhibition networks could imply that self-control involves two stopping systems that rely on external components and internal components respectively, and that the degree of their interplay depends on the individual's motivational state. In 'cold' or urge-free situations, the two systems operate in isolation and are largely distinguishable. In this context, intentional inhibition reflects primarily the endogenous decision to cancel an ongoing behavior, and it is less clear to what extent such inhibition reflects genuine motor suppression or rather higher-order disengagement (see Lynn et al., 2014, for a theoretical account of the latter view). Externally guided inhibition, on the other hand, assesses the rapid implementation of behavioral suppression, absent decisional aspects, as the response outcome is fully determined by the environment. However, given the presence of a strong behavioral urge, such as the desire to avoid physical pain, the two systems must necessarily interact to enable successful suppression of the urge, regardless of the level of instruction. In this context, both the decisional and the implementation aspects of behavioral inhibition must continuously be reinforced in order to adhere to the circumstantial demands. Future research will be necessary to test this idea more directly. For instance, comparing intentional and externally guided inhibition in both 'hot' and 'cold' response systems within the same subjects could provide direct evidence for the aforementioned view.

Conclusion

The present study investigated the neural basis of the intentional inhibition of pain avoidance behavior. Replicating previous work, we found the dFMC, rIFG extending into the insula, and pre-SMA to exhibit activity related to suppression of the pain withdrawal urge, when compared with executing an abating response. However, this pattern was observed regardless of the locus of the inhibit decision, suggesting that the degree of intentional engagement depends on one's motivational state. Our paradigm therefore provides a novel and ecologically-valid context for the investigation

of intentional inhibition, but future studies will be needed to determine whether everyday instances of exogenous and endogenous inhibition are dissociable.

Acknowledgements

We thank Pieter Vandemaele for technical support and Paul S. Muhle-Karbe for helpful comments on the manuscript. This work was supported by the European Science Foundation's EUROVETO project (09-ECRP-020).

REFERENCES

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience*, *26*, 2424–2433.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170–7.

Aron, A.R., Robbins, T.W., & Poldrack, R.A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences, 18*(4), 177 - 185.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355.

Brass, M., Derrfuss, J., & von Cramon, D.Y. (2005). The inhibition of imitative and overlearned repsonses: A functional double dissociation. *Neuropsychologia, 43,* 89 - 98.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *The Journal of Neuroscience*, *27*(34), 9141–5.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist*, *14*, 319–325.

Brass, M., & Haggard, P. (2010). The hidden side of intentional action: the role of the anterior insular cortex. *Brain Structure & Function*, *214*, 603–610.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behavior and social cognition. *Philosophical Transactions of the Royal Society of Biological Sciences, 364,* 2359 - 2367.

Brett, M., Anton, J.-L. L., Valabregue, R., & Poline, J.-B. B. (2002). Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. In *NeuroImage* (Vol. 16, p. abstract 497).

Brody, A. L., Mandelkern, M. A., Olmstead, R. E., Jou, J., Tiongson, E., Allen, V., … Cohen, M. S. (2007). Neural substrates of resisting craving during cigarette cue exposure. *Biological Psychiatry*, *62*, 642–651.

Cai, W., George, J. S., Verbruggen, F., Chambers, C. D., & Aron, A. R. (2012). The role of the right presupplementary motor area in stopping action: two studies with event-

related transcranial magnetic stimulation. *Journal of Neurophysiology*, *108*(2), 380–9.

Campbell, B.A. & Misanin, J.R. (1969). Basic drives. *Annual Review of Psychology, 20,* 57 - 84.

Campbell-Meiklejohn, D. K., Woolrich, M. W., Passingham, R. E., & Rogers, R. D. (2008). Knowing when to stop: the brain mechanisms of chasing losses. *Biological Psychiatry*, *63*, 293–300.

Chikazoe, J., Jimura, K., Hirose, S., Yamashita, K., Miyashita, Y., & Konishi, S. (2009). Preparation to inhibit a response complements response inhibition during performance of a stop-signal task. *The Journal of Neuroscience*, *29*, 15870–15877.

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activation of right inferior frontal gyrus during response inhibition across response modalities. *Journal of Cognitive Neuroscience*, *19*, 69–80.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*, 1325–1335.

Elliot, A. J. (2006). The Hierarchical Model of Approach-Avoidance Motivation. *Motivation and Emotion*, *30*(2), 111–116.

Filevich, E., Kühn, S., & Haggard, P. (2012). Intentional inhibition in human action: the power of “no”. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1107–18.

Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2008). Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: a model-based approach. *The Journal of Neuroscience*, *28*, 9790–9796.

Garavan, H., Ross, T. J., Murphy, K., Roche, A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection and correction. *NeuroImage*, *17*, 1820–1829.

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *PNAS*, *96*, 8301–8306.

Garcia-Larrea, L., & Peyron, R. (2013). Pain matrices and neuropathic pain matrices: a review. *Pain*, *154 Suppl 1*, S29–43.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews. Neuroscience*, *9*(12), 934–46.

Herz, D.M., Christensen, M.S., Bruggemann, N., Hulme, O.J., Ridderinkhof, K.R., Madsen, K.H., & Siebner, H.R. (2014). Motivational tuning of fronto-subthalamic connectivity facilitates control of action impulses. *The Journal of Neuroscience, 34(9),* 3210 - 3217.

Hodgson, T., Chamberlain, M., Parris, B., James, M., Gutowski, N., Husain, M., & Kennard, C. (2007). The role of the ventrolateral frontal cortex in inhibitory oculomotor control. *Brain : A Journal of Neurology*, *130*, 1525–1537.

Ingvar, M. (1999). Pain and functional imaging. *Philosophical Transactions of the Royal Society of London*, *354*(1387), 1347–58.

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. *PNAS*, *101*(36), 13335–40.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Touryan, S. R., Greene, E. J., & Nolen-Hoeksema, S. (2006). Dissociating medial frontal and posterior cingulate activity during self-reflection. *Social Cognitive and Affective Neuroscience*, *1*, 56–64.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–40.

Kühn, S., Gallinat, J., & Brass, M. (2011). “Keep calm and carry on”: structural correlates of expressive suppression of emotions. *PloS One*, *6*, e16569.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: how the “veto-area” exerts control. *Human Brain Mapping*, *30*(9), 2834–43.

Lynn, M.T., Muhle-Karbe, P.S., & Brass, M. (2014). Controlling the self: The role of the dorsal frontomedian cortex in intentional inhibition. *Neuropsychologia.*

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, *106*(1), 3–19.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., … Shoda, Y. (2011). “Willpower” over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*(2), 252–6.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 17, 1306–1315.

Morsella, E. (2005). The function of phenomenal states: supramodular interaction theory. *Psychological Review*, *112*(4), 1000–21.

Mueller, V., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, *37*(4), 1354–61.

Muraven, M. (2010). Practicing self-control lowers the risk of smoking lapse. *Psychol Addict Behav*, *24*(3), 446–452.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin*, *126*(2), 247–59.

Nachev. P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience, 9,* 856 - 869.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Peyron, R., Laurent, B., & García-Larrea, L. (2000). Functional imaging of brain responses to pain: A review and meta-analysis. *Clinical Neurophysiology*, *30*(5), 263–88.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, *277*, 968–971.

Swann, N.C., Cai, W., Conner, C.R., Pieters, T.A., Claffey, M.P., George, J.S., Aron, A.R., & Tandon, N. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage, 59*(3), 2860 - 2870.

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.

Towse, J. N. & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers, 30 (4),* 583 - 591.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418–424.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., … Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science, 303,* 1162–1167.

CHAPTER 2

Controlling the self: The role of the dorsal frontomedian cortex in intentional inhibition[1]

Intentional inhibition refers to the suppression of ongoing behavior on the basis of internally-generated decisions. This ability to cancel planned actions at the last moment is thought to be critical for self-control and has been related to activation in a circumscribed region of the dorsal frontomedian cortex (dFMC). Preliminary theories of intentional inhibition were based on studies that exclusively examined the cancellation of motor responses, and consequently concluded that this region serves the suppression of motor output. Yet recent evidence suggests that the dFMC is also involved in inhibitory control over more abstract internal states such as emotions or desires that have no immediate behavioral output. In this review, we therefore wish to put forth a new integrative perspective on the role of the dFMC in human self-control. We will argue that by virtue of its anatomical location and functional connections, this area may subserve the disengagement from current urges and impulses, thus facilitating successful exertions of self-control across a wide range of contexts by overcoming a self-focused perspective. We will discuss the fit of this view of the dFMC with the existing literature, identify critical experimental determinants for engaging the dFMC in intentional inhibition, and outline promising perspectives for future research.

[1] Lynn, M.T., Muhle-Karbe, P.S., & Brass, M. (2014). Controlling the self: The role of the dorsal frontomedian cortex in intentional inhibition. *Neuropsychologia*.
doi: 10.1016/j.neuropsychologia.2014.09.009

INTRODUCTION

The ability to withhold behavioral impulses in favor of higher-order goals is central to human self-regulatory behavior (Baumeister et al., 2007). To date, this ability has been investigated in two distinct research domains, namely cognitive and social psychology. In cognitive psychology, research on inhibitory control typically employs experimental paradigms that require participants to withhold simple key presses in response to pre-instructed stop or nogo signals (henceforth referred to as externally guided inhibition, Chikazoe et al., 2007; Logan & Cowan, 1984; Verbruggen & Logan, 2008, 2009). While this research has the advantage of providing maximal experimental control, the transfer to inhibition and self-control in everyday life is limited. In particular, most situations that require self-control do not provide external signals that indicate whether or not to withhold a specific behavior. By contrast, social psychological research on self-control has investigated behavioral inhibition under more complex and ecologically valid circumstances (see Hagger et al., 2010 for a review). Because of the complex experimental settings that are used, however, it is often very difficult to determine the exact neurocognitive mechanisms that are involved in these forms of self-control. Recently, the theoretical concept of *intentional inhibition* has been introduced to combine elements of both research traditions (Brass & Haggard, 2007; Brass & Haggard, 2008; Filevich et al., 2012). Intentional inhibition refers to the voluntary and internal decision to withhold from executing a prepotent action tendency. In this sense, the concept of intentional inhibition is much closer to social psychological conceptions of self-control.

In the current review, we will first give an overview of research in the domain of intentional inhibition. We will argue that the concept strongly relies on the assumption that intentional inhibition can be distinguished from externally guided inhibition on the basis of its functional neuroanatomy. In particular, intentional inhibition has been related to a specific part of the medial prefrontal cortex, namely the dorsal frontomedian cortex (dFMC), although the precise functional contribution of this region remains elusive. Therefore we will try to explore the role of the dFMC in the broader context of self-control. Thereafter we will argue, based on the location of this area at the intersection of brain areas involved in cognitive motor control and those involved in more complex self-reflective and social cognitive processes, that it contributes to self-

control by facilitating disengagement from impulses and urges[2]. On the basis of this new conception, we will outline crucial experimental determinants for investigating intentional inhibition and sketch future perspectives in this research domain.

Previous findings: What do we know about intentional inhibition?

The concept of intentional inhibition is relatively young and only dates back a few years. As outlined above, it evolved as an extension of classical inhibition research in cognitive psychology, which focused primarily on externally guided inhibition. Following the logic of intentional action research, in which intentional action is usually contrasted with stimulus-guided action (e.g., Passingham et al., 2000; Toni et al., 2001), early paradigms tried to introduce a choice component preceding the inhibition process. From this perspective, intentional inhibition, like intentional action, is internally generated. Yet because intentional inhibition paradigms do not present a stop signal, it becomes difficult to derive chronometric measures such as stop signal reaction times (SSRTs). Thus, the only behavioral dependent measure that can be used is the proportion of inhibition trials (e.g. Brass & Haggard, 2007; Lynn et al., 2013; Rigoni et al., 2012). Brain imaging techniques such as fMRI therefore provide valuable tools that permit the delineation of neural activity preceding intentional decisions to inhibit behavior.

In the first study to introduce the concept of intentional inhibition, Brass & Haggard (2007) employed a modified version of the method introduced by Libet et al. (1983), which required participants to perform self-paced button presses (i.e., action trials) and to monitor the moment in time when they felt the intention to execute the movement. In addition, participants were instructed to occasionally prepare such movements but cancel them at the very last moment prior to execution (i.e., inhibition trials). Contrasting brain activity on inhibition trials with action trials yielded increased activation in the dFMC, as well as in the left and right anterior insula, and the superior temporal sulcus. This neural signature was in striking contrast to findings resulting from externally guided response inhibition, which typically engages a neural network around the right inferior frontal gyrus, the pre-supplementary motor area (pre-SMA), and the

[2] We use the term 'impulse' to denote particularly sudden or spontaneous response tendencies, whereas 'urge' refers to a subjective motivation that develops over time.

basal ganglia (rIFG; see Aron et al., 2004; 2014 for reviews), implying that these two types of behavioral inhibition rely on largely different control mechanisms.

Importantly, a number of follow-up studies that employed quite disparate experimental procedures could replicate the involvement of the dFMC in intentional inhibition. For instance, Kühn et al. (2009) introduced the so-called "ramp task" in which participants saw the image of a marble moving downwards on a ramp and breaking into pieces when it reached the end of the ramp. Participants could freely choose between preventing the marble from breaking via a button press, and inhibiting their urge to do so. Importantly, the shattering of the marble was associated with an aversive glass-breaking sound and a monetary loss in order to create an incentive of responses over response omissions, which ties intentional inhibition more closely to realistic scenarios of self-control in which highly prepotent behavior needs to be suppressed. As in the study by Brass & Haggard (2007), inhibition-related brain activity was found in the dFMC, leading to the idea that the dFMC constitutes a "veto area" that generates endogenous top-down signals in the service of the intentional cancellation of behavior. However, this interpretation of the dFMC as being involved in intentional inhibition rests on the assumption that activity in this brain area precedes the inhibition process. Given that the temporal resolution of fMRI is rather poor, this assumption requires independent proof. A recent EEG study addressed this question (Walsh et al., 2010) using a variant of the Libet task similar to Brass & Haggard (2007). Frequency analyses of brain oscillations shortly after the experience of an intention to move revealed an increase in spectral power over frontal electrodes that was specific for trials in which the movement was then inhibited, highlighting that the neural signature of intentional inhibition has a plausible time course during motor preparation.

The role of the dFMC in intentional inhibition

Despite this converging evidence for the general importance of the dFMC in intentional inhibition, it is still largely unknown *how* this area exerts control over behavioral impulses. Recently, Filevich et al. (2012) embedded the concept of intentional inhibition in a more general model of motor control, based on internal feedback loops. This model includes an inner loop that continuously adjusts movement parameters based on a comparison between predicted and perceived sensory feedback,

and an outer loop that monitors the long-term consequences of ongoing actions and compares them with general goals. Intentional inhibition is conceived as a braking mechanism that links both loops by cancelling ongoing behavior when the anticipated outcome seems no longer desirable. In line with this idea, the dFMC has been shown to exhibit increased functional connectivity with the pre-SMA during intentional inhibition (Kühn et al., 2009; 2013). This finding is of particular interest, given that the pre-SMA is involved in action planning (e.g., Cunnington et al., 2002; 2003) as well as in externally guided response inhibition (e.g., Simmonds et al., 2008). Accordingly, the pre-SMA may constitute a common pathway for the implementation of different types of motor decisions, with the dFMC directing the outcome of this decision.

Additional research has concurrently shown that the dFMC is not only involved in the inhibition of overt behavior, but also in the suppression of other psychological states such as emotions (Kühn et al., 2011; 2013), cigarette cravings (Brody et al., 2007; Hanlon et al., 2013; Hartwell et al., 2012), or gambling desires (Campbell-Meiklejohn et al., 2008). While these findings strengthen the general notion that the dFMC is a crucial brain region for the successful exertion of self-control, they also indicate that the functional contribution of this region might be more general than previously assumed, and extend beyond the suppression of motor output.

In the following, we will put forth a broader perspective of the role of the dFMC in self-control, arguing that this area allows for disengagement from one's current impulses and urges. We will first review literature from social psychology to highlight that such disengagement strategies are an effective functional mechanism for exerting self-control. Thereafter we will outline that the dFMC is well situated to perform this self-regulatory function because of its anatomical location and functional connections. Finally, we will illustrate that this new conception of intentional inhibition is capable of explaining seemingly inconsistent findings in the literature, and helps to further bridge the gap between cognitive and social psychological conceptions of inhibition.

Disengagement as a functional mechanism for self-control

Evidence for the role of disengagement in self-control dates back to the famous delay-of-gratification experiments (e.g., Mischel, 1974) in which preschoolers were able to choose between a smaller but immediately available reward, and a larger reward

with a temporal delay. Successful delay strategies on the part of the children included self-distraction or redirection of attention (e.g., hiding their faces) and altering the way they mentally represented the desired object (i.e., reappraisal or reframing; see Mischel et al., 2011 for a recent review). Interestingly, such distancing strategies do not appear to be used exclusively by children, but persist into adulthood. For example, while engaged in a smoking cessation program, smokers tend to initially psychologically distance themselves from their archetype of the typical smoker, until such time as they have successfully broken their habit (Gibbons et al., 1991). Likewise, in the realm of emotion suppression, it is well known that people employ specific reappraisal strategies, most notably distancing, when attempting to decrease their emotional response to stimuli (e.g., Gross, 1998; 1999). Indeed, participants in several emotion suppression studies were specifically trained and instructed to maintain an objective viewpoint by detaching themselves from the emotional target stimulus during inhibition trials (Kühn et al., 2013; Lévesque et al., 2003; Ochsner et al., 2004).

Conversely, when the capacity for self-control is impaired, as in the case of ego depletion (Baumeister et al., 1998), people tend to adopt a self-focused, first-person perspective. In the case of Achtziger et al. (2011), ego depleted participants engaged in an Ultimatum game employed maladaptive and self-centered strategies in which they not only proposed lower amounts for their partner, but rejected low offers themselves, accepting instead zero financial reward for both players, rather than attending to an absolute monetary calculation or value. Similarly, Macrae et al. (2014) found that self-regulatory failures were exacerbated both when adopting a first- versus third-person perspective, and when adopting a near versus far distal perspective (under both temporal and physical conditions).

Altogether, these findings indicate that disengagement from one's current urges and impulses is indeed an effective means of boosting one's regulatory abilities. This idea is furthermore in line with recent theoretical work on self-control by Fujita (2011), who describes it as "the general process by which people advance abstract, distal over concrete, proximal motives in judgment, decisions, and behavior." Accordingly, to successfully self-regulate, one must achieve a measure of mental remove from their present circumstances.

The dFMC is well situated to support disengagement from urges

Having illustrated that distancing is an adaptive self-regulatory strategy, this section will describe why the dFMC is a brain region particularly well situated to perform this function. The past two decades of brain imaging research have yielded a general consensus that endogenous control processes rely most critically upon the integrity of the medial frontal cortex (MFC; see Amodio & Frith, 2006; Passingham et al., 2010; Rushworth et al., 2004 for reviews). At the same time, the MFC does not constitute a functional unit, but rather consists of distinct sub-regions that serve dissociable functions (see Fig. 1). As such, the posterior parts of the MFC are clearly implicated in cognitive motor control (Krieghoff et al., 2011; Nachev et al., 2008). Two regions have been of particular interest in this respect: the rostral cingulate zone (RCZ), a region extending dorsally and caudally from the anterior cingulate cortex (ACC), and the supplementary and pre-supplementary motor areas (SMA/pre-SMA). The RCZ has been identified in disparate lines of research and linked with both resolving conflict between incompatible response tendencies (Botvinick et al., 2004; Lau et al., 2004;2006; Ridderinkhof et al., 2004), and in the intentional selection of movements (Krieghoff et al., 2009; Müller et al., 2007; Walton et al., 2004) and tasks (Demanet et al., 2013; Forstmann et al., 2006). Brass & Haggard (2008) argued that the RCZ's sensitivity to both conflict and volition might reflect a common process of choosing between different response alternatives. This integrative perspective has recently been formalized in a computational model (Holroyd & Yeung, 2012). The SMA/pre-SMA is also critical for the planning and control of movements, yet its contribution seems to be more closely related to the implementation of motor decisions. On the one hand, this area has been linked to the planning and generation of self-initiated movements (Cunnington et al., 2002; 2003; Debaere et al., 2003; Deiber et al., 1999). In line with this, electrical stimulation of the pre-SMA can induce the sensation of an "urge to move" (Fried et al., 1991). At the same time, the SMA/pre-SMA is also critical for re-programming and inhibiting movements (Nachev et al., 2008; Neubert et al., 2010; Sumner et al., 2007). This is also illustrated by the so-called "anarchic hand syndrome" that can result from unilateral lesions to the pre-SMA. Patients with anarchic hand syndrome exhibit an inability to inhibit stimulus-induced action tendencies with the hand contralateral to the lesion, despite the phenomenological experience of having an

intention to do so (Della Sala et al., 1991; Kritikos et al., 2005; Pacherie, 2007). Accordingly, the SMA/pre-SMA may be critical for adjusting the degree of action readiness (Cunnington et al., 2005; Forstmann et al., 2008) that can yield facilitation or suppression of motor output, presumably via projections to the basal ganglia and the primary motor cortex (Picard & Strick, 2001). It is currently unclear to what extent the SMA/pre-SMA itself is the generator of top-down motor control, and to what extent it relies on input from other areas, such as the dFMC or the rIFG.

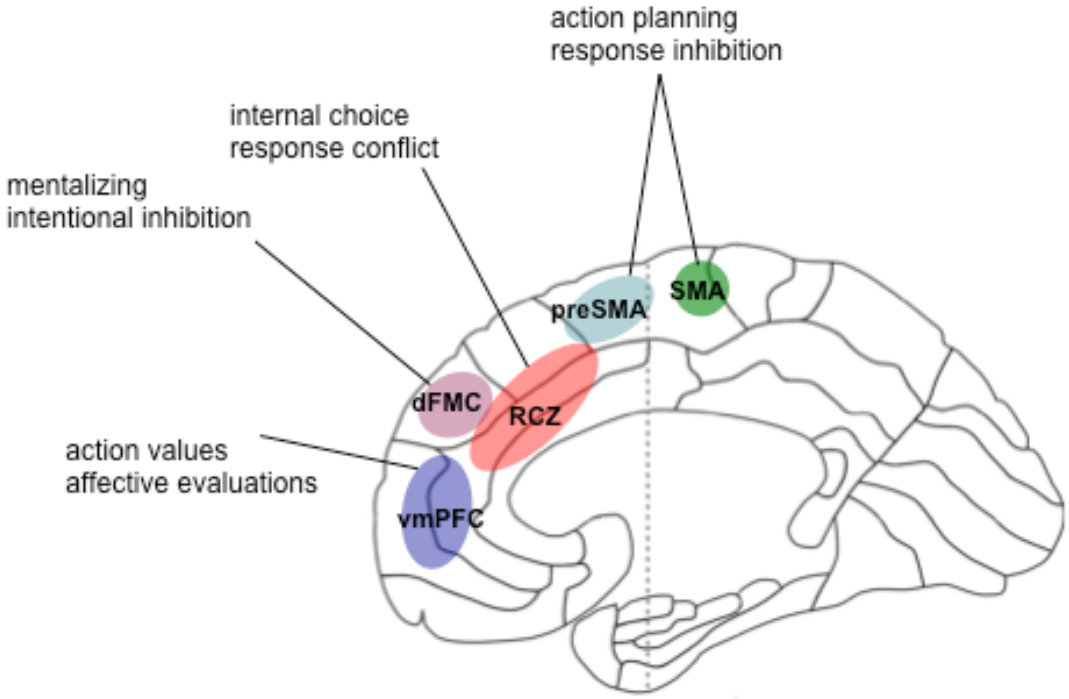

Figure 1. ***Schematic overview of brain regions in the medial frontal wall, along with their assumed functions.***

In contrast to these motor control functions of the posterior MFC, the anterior MFC is involved in higher-order self-referential, social-cognitive, and emotional processes. The anterior MFC is typically divided into a ventral part and a dorsal part (Amodio & Frith, 2006; Forbes & Grafman, 2010; Lieberman, 2007; Mitchell et al., 2006; Van Overwalle, 2009). The ventral part has been associated with automatic affective evaluations and the computation of reward values in decision making (Bechara et al., 1999; Bechara, 2005; Kolling et al., 2012; Mauss et al., 2007; Reuter et al., 2005; van den Bos et al., 2007). The dorsal part, on the other hand, has been linked with both

mentalizing about other people's internal states (Frith & Frith, 2003; Gallagher & Frith, 2003; Gilbert et al., 2006; Zaki et al., 2009) and active self-referential thought processes (Andrews-Hanna et al., 2010; Whitfield-Gabrieli et al., 2011). This raises the question of how the very abstract role of the dFMC in social cognition can be reconciled with its role in intentional inhibition (see Fig. 2 for an illustration of the overlap). Although it is possible that different functions co-exist within the dFMC, we believe that intentional inhibition and mental state attribution share a common functional mechanism; both require disengagement from a strong self-perspective. In mental state attribution, overcoming the self-perspective is needed to attribute mental states to others that are inconsistent with one's own perspective. In intentional inhibition and self-control, overcoming the self-perspective is necessary for successfully disengaging from a strong impulse or urge to act.

Within the what-when-whether (WWW) model of intentional action (Brass & Haggard, 2008; Brass et al., 2013), the dFMC would serve a regulatory function between areas that are involved in intentional choice (RCZ) and areas involved in the implementation of such choices (pre-SMA). The dFMC would enable the overcoming of a strong impulse or urge by down-regulating brain areas that are involved in the formation and implementation of such impulses, leading to a disengagement from the self-perspective. This is consistent with the functional connectivity data of the dFMC and pre-SMA outlined above. Furthermore, from this perspective, the frequent co-activation of the dFMC with the anterior insula in intentional inhibition (e.g., Brass & Haggard, 2007; Campbell-Meiklejohn et al., 2008; Lynn et al., in revision) might be re-evaluated. Thus far, the involvement of the insula has primarily been linked to a "feeling of let down" that may result from negative evaluations of not carrying out a response after preparing it (see Brass & Haggard, 2007). However, the prominent role of this region in awareness (see Craig, 2009 for a review) could also imply that the role of the insula in intentional inhibition is self-reflective in nature.

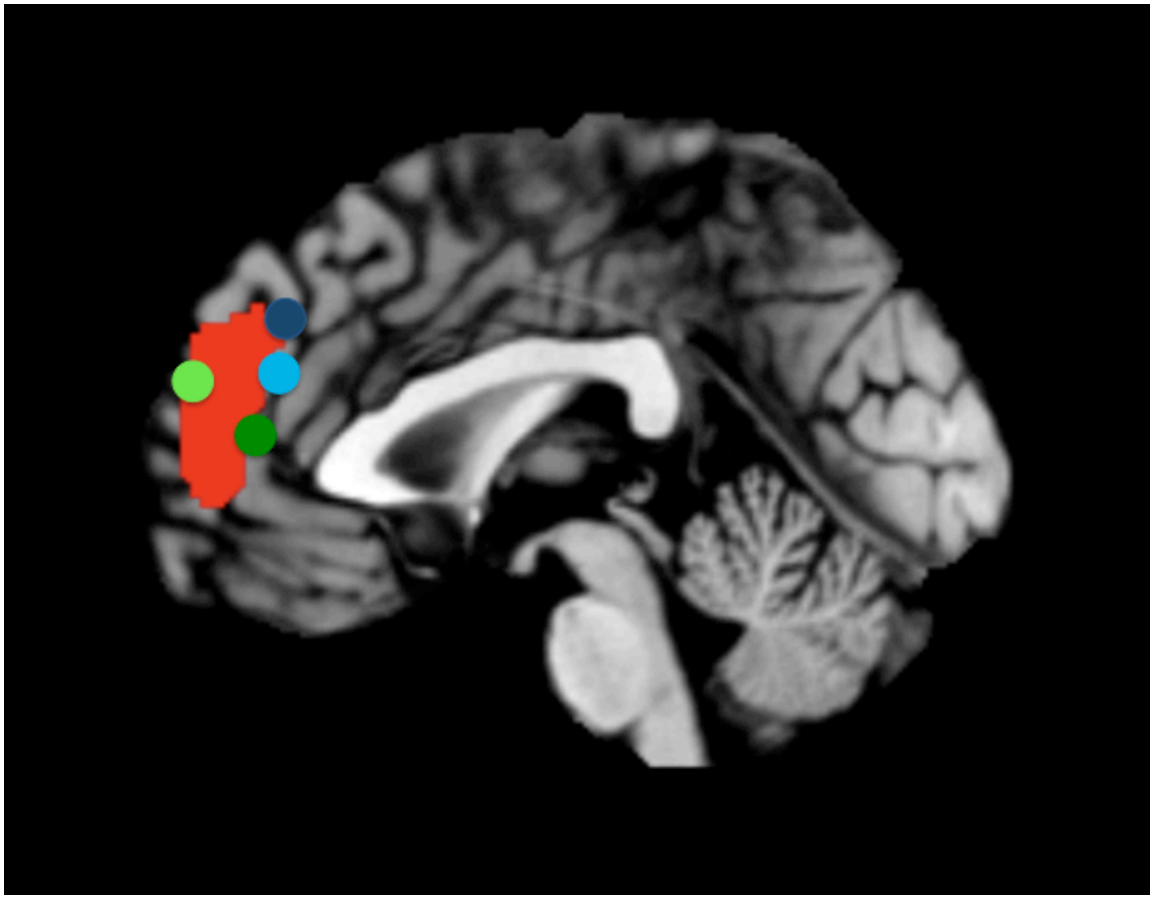

Figure 2. ***Overlap of brain activation related to mentalizing and intentional inhibition. The red cluster represents the results of an activation-likelihood estimation meta-analysis of 38 theory of mind studies (including 47 peaks) taken from Amodio & Frith, 2006 and Gilbert et al., 2006. Peaks from four intentional inhibition studies are represented as dots. Dark blue = Brass & Haggard 2007; light blue = Kühn et al., 2009; light green = Kühn et al., 2013; dark green = Lynn et al., in revision. The former two studies were motoric in nature, while the latter two involved emotions and urges, respectively.***

Critical determinants for engaging the dFMC

We will next revisit existing literature on intentional inhibition and evaluate to what extent the disengagement account is capable of explaining the data. Intentional inhibition has been applied to various contexts and experimental procedures, resulting in a heterogeneous pattern of findings with some studies replicating the involvement of the dFMC and others not. Based on the idea that this region supports the process of disengagement, we will first propose crucial determinants that experimental paradigms might need to satisfy in order to tap into this self-regulatory function. To elicit intentional disengagement it seems critical that the experimental setting meets three principle demands: First, participants must be operating under the circumstances of

choice, and have to make an explicit decision to stop the behavior. Second, there must be enough *time* in which to form and implement this decision. Otherwise participants might decide to act or inhibit in advance of the trial (pre-decision) or, conversely, generally omit responding first and thereafter decide (post-decision). Finally, the decision to inhibit must be imposed in the context of a strong *urge or impulse* to act. Choice, temporal considerations, and urges may also interact; for example, the strength of a behavioral urge can modulate the impact of an internal choice, as in the case of holding one's breath. The body will permit oxygen starvation only up to a point, after which the urge to inhale will overcome all other response options. Similarly, choice can be implicitly biased via reward incentives. Below we will revisit existing literature on intentional inhibition and examine to what extent these paradigms have met the hypothesized determinants.

Paradigms that identified dFMC activity with intentional inhibition

So far, three studies that were explicitly designed to investigate intentional inhibition have revealed inhibition-related activity in the dFMC. As described above, the study by Brass & Haggard (2007) employed a variant of the Libet task, a paradigm that clearly provides both a free choice between acting and not acting, and sufficient time to make this decision. Importantly, participants were asked to monitor the moment in time when they felt the intention to press the button. This likely increased the self-reflective engagement in the task and discouraged participants from pre-deciding to act or inhibit in advance of the trial. The only caveat of this paradigm resides in the fact that the choice between acting and not acting is relatively arbitrary, since there is no genuine incentive for either option. Nevertheless, dFMC activity was found when contrasting omissions versus action trials, located in the medial frontal gyrus dorsal to the anterior cingulate.

The second intentional inhibition paradigm, the ramp-task, subsequently introduced by Kühn et al. (2009), sought to overcome the problem of arbitrary response options and created a prepotency of acting over inhibiting. This was achieved by linking actions to instrumental outcomes (i.e., the avoidance of the aversive marble breaking sound, and a trial-based monetary reward). In this sense, the ramp task was a clear step towards situating intentional inhibition in more realistic settings by inducing a strong

impulse to act. Moreover, this paradigm provided a free choice between acting and inhibiting. Consistent with the aforementioned determinants, dFMC activity was found for intentional inhibition, located slightly more ventrally than in the study by Brass & Haggard (2007). However, this paradigm has a number of limitations compared to the Libet task; the timing of the decision to act or inhibit is linked to the external event of the marble beginning to move, and participants were not required to monitor their internal states (e.g., their intention/urge to press the button), thus making pre-decisions on the participants' part more likely. Finally, the paradigm introduced substantial time pressure on the decision to inhibit or to act. These aspects may explain why the original findings resulting from this task have proven somewhat difficult to replicate absent reward. Schel et al. (2014) compared the ramp task to a stop-signal task, and found dFMC activation in the former task not for intentional inhibition versus stimulus-driven action, as in the previous study, but in a parametric analysis of the number of preceding go trials. Here, stronger dFMC activity (at a location similar to that of Kühn et al., 2009), a higher proportion of inhibition trials, and slower RTs were observed for choice trials when there were fewer preceding go trials. One interpretation (supported by the reaction times in particular) could be that responding becomes more automated following long sequences of instructed trials, and the more demanding decisions arise when alternating rapidly between instructed and free-choice scenarios, or when trying to keep one's decisions random in an extended choice scenario.

Another context to which intentional inhibition has been applied is emotion regulation (Kühn et al., 2013). Highly arousing images with negative valence were presented for a duration of two seconds before a cue appeared that either instructed participants to inhibit their emotional response to these images by means of distancing (exogenous inhibit), to allow their feelings to unfold (exogenous feel), or to chose between one of the two options (endogenous inhibit, endogenous feel). Thus, in contrast with previous emotion regulation literature, this study provided a condition that allowed participants to engage in emotion regulation on the basis of free choice. Moreover, participants were asked to rate the intensity of the emotions that they felt on each trial, making pre-decision strategies rather unlikely. Accordingly, the dFMC activation, observed for endogenous, but not exogenous, emotion inhibition is

consistent with our proposal. Here, the focus of activation was located more rostrally than in previous studies, clearly beyond the cingulate gyrus.

Studies that identified dFMC activity with externally guided inhibition

Interestingly, there are also a few studies that have found dFMC activation for externally guided inhibition. We presume that this is due to the fact that, although participants were cued to inhibit, the urge to act in these paradigms was so strong that even in the presence of a stop signal they had to explicitly decide to inhibit. The first study tried to introduce a more ecologically valid experimental setting by examining the suppression of an extremely strong behavioral urge, namely the avoidance of pain (Lynn et al., in revision). Here, participants received thermal pain stimulation to alternating inner wrists and could terminate the stimulation via button presses with the opposing hand. The benefit of this study was that the urge to inhibit built organically over time, and was present in each trial. In addition, participants were given sufficient time to make a decision on choice trials, with an equivalent proportion of choice and directed trials, discouraging them from pre-deciding. Accordingly, this paradigm could be expected to elicit a strong urge to act and to trigger disengagement in order to successfully inhibit the urge. Indeed, dFMC activation was observed when participants inhibited their pain avoidance response (peaking more ventrally in the amPFC but extending dorsally into the dFMC). However, in this study, equivalent dFMC activity was observed for both intentional and externally guided inhibition. This may reflect that the suppression of very strong urges relies on intentional disengagement regardless of how the decision to inhibit is initially determined.

This interpretation is bolstered by similar findings in the realm of addiction research, particularly the suppression of cigarette cravings (Brody et al., 2007, Hanlon et al., 2013, Hartwell et al., 2013). In these studies, participants were instructed to apply different strategies when faced with craving-inducing stimuli, namely either to resist the craving via self-distraction or other strategies, or to permit the craving sensation to unfold. Although the response was entirely determined by the instruction, reliable activity was found in large clusters along the dFMC and the anterior cingulate when contrasting resist with permit conditions. Accordingly, when faced with extremely strong behavioral urges (particularly those involving self-preservation or physical

needs), the difference between intentional and externally guided inhibition seems to be diminished because the demand to override prepotent response tendencies requires strong activation of the goal to inhibit. Instruction alone is insufficient to achieve that end, so internal resources must be recruited in order to comply with the experimental constraints.

Studies that did not identify dFMC activity with intentional inhibition

Finally, we would like to also mention two studies that did not identify inhibition-related dFMC activity, despite superficially appearing to involve similar demands. The first study was designed to mimic a stop signal paradigm but with a choice cue that left it open to participants whether to inhibit the response or not (Kühn & Brass, 2009). In 75% of trials, participants were instructed to press one of four keys in response to a stimulus in order to induce a very strong response tendency. When the color of the stimulus changed they either had to withhold the key press or to decide to press or inhibit depending on the color. When comparing trials where participants decided to inhibit (decide nogo) with trials where a stimulus indicated not to act (instructed nogo), massive activation in the RCZ was found, but no activation in dFMC. When comparing the decide nogo with the instructed nogo, no activation was found at all. SSRTs for instructed nogo and decide nogo were very similar, indicating that in a situation where the time pressure to decide between inhibiting and acting is very high, participants first inhibit the response and then decide whether or not to reinitiate the response. This later decision, however, does not require any disengagement from a strong impulse because the action had been inhibited beforehand. Therefore, no activation of the dFMC could in fact be expected.

Another study that tried to investigate intentional inhibition in a more ecologically-valid context examined the inhibition of taboo word utterances (Severens et al., 2012). Participants completed a word reading task with stimulus pairs that frequently elicit spoonerisms. For half of the word pairs, spoonerisms resulted in the utterance of socially inappropriate taboo words, and for the other half in neutral expressions. The authors contrasted taboo with neutral trials, under the assumption that taboo word pairs would elicit additional inhibition processes in order to avoid the utterance of socially inappropriate expressions. This contrast yielded activity in the right

IFG, but not in the dFMC. Thus, although this paradigm did not entail any explicit external stop signals, the observed brain activity was very similar to studies investigating externally guided inhibition. This could be related to the fact that there was no real choice for participants in this study (both the explicit instruction to read word pairs and the implicit instruction to inhibit spoonerisms were constant across all trials). In addition, the speeded nature of the task did not leave any time for decision-making. Finally, it needs to be considered that in this case a social norm (not to utter socially undesirable expressions) might be internalized and act as a stop signal.

To summarize, it appears rather difficult to define a priori properties of tasks that can be used to predict dFMC activation in intentional inhibition, yet the criteria based on the disengagement account appear to provide a stable means of navigating existing data. Paradigms that have been successful in revealing dFMC activation balance different characteristics. First, as outlined above, inhibition must be based on a decision to disengage from a strong impulse or urge. However, such a decision does not necessarily have to be an explicit requirement of the experimental paradigm but can also be induced by a very strong impulse to act. In situations where the urge to act is extremely strong, participants might be involved in a decision process even when they are explicitly told to inhibit and therefore formally do not have to decide. On the other hand, in situations where no explicit instruction to inhibit is given, inhibition might be nevertheless automatically induced by contextual information affording inhibition of the action (e.g. the inhibition of taboo words). Secondly, the timing of the inhibition process is crucial, with too much time pressure leading to ‘offline’ decisions to inhibit. However, if participants are not engaged in a specific behavior at the moment they decide to inhibit, no disengagement is required.

Given the above critical determinants of disengagement, and taking into account the varied paradigms employed thus far to investigate intentional inhibition, one may directly assess the intersection of self-control strategies and intentional inhibition by optimizing experimental designs to more closely test the disengagement hypothesis. It would also be useful to manipulate the disengagement strategy itself, by inducing a first- or third-person perspective prior to an intentional inhibition task that conforms to the above constraints. In this case, one might find that, in comparison to a

first-person perspective, an objective stance promotes both behaviorally measured self-control, and dFMC activation.

Distinguishing intentional inhibition from proactive inhibition

Finally, we would like to distinguish the concept of intentional inhibition from the related concept of proactive inhibition. Recently, proactive inhibition has been introduced in an attempt to relate stop-signal response inhibition more closely to impulse control in realistic scenarios (see Aron, 2011 for a review). To this end, studies have examined how participants adjust their motor preparation when advance information is given about either the likelihood of a stop signal (e.g., by comparing blocks with and without stop signals) or its specificity (e.g., by comparing stop signals that apply to all vs. only certain possible responses). Accordingly, this setup allows for the investigation of how one "proactively" adjusts their response preparation for the possibility of a stop signal, rather than only focusing on the "reactive" implementation of inhibition in response to it. Preliminary evidence resulting from this procedure indicates that participants are able use the advance information provided by the cues (i.e., their SSRTs on nogo trials is reduced and their RTs on go-trials is increased when they know a stop may occur), and these adjustments seem to rely on similar brain areas as the reactive implementation of inhibition (Chikazoe et al., 2009; Jahfari et al., 2010; Zandbelt et al., 2013). This underscores the notion that the external inhibition network (i.e., the rIFG, the pre-SMA, and the basal ganglia) is not only activated bottom-up by the appearance of a stop signal, but also top-down on the basis of strategic adjustments.

However, although proactive inhibition can be considered as resulting from endogenous preparation processes, it still differs from the scope of intentional inhibition as outlined in this review. Most importantly, proactive inhibition is still concerned with stimulus-guided behavior in which the outcome (action or omission thereof) will be determined by external events, whereas intentional inhibition refers to the cancellation of behavior on the basis of free choice and internal states. In addition, proactive inhibition influences primarily the readiness to perform or inhibit an action, whereas intentional inhibition operates on ongoing behavior. Nevertheless, it could be an interesting perspective for future research to examine how the brain networks

involved in proactive and intentional inhibition may interact, especially under motivationally salient circumstances.

CONCLUSION

In this review, we have summarized previous literature on intentional inhibition, highlighting that this capacity has been frequently linked with the dFMC, though the precise role of this region has been elusive. We propose a disengagement hypothesis, whereby the dFMC contributes to self-control by separating a person from their immediate urges. This assumption is consistent with the location of the dFMC at the interface of motor control and social cognition. Furthermore, this hypothesis provides the means for an integrative account of intentional inhibition, moving beyond motor control and encompassing social psychological conceptions of self-regulation. Attending to the critical determinants of intentional inhibition would permit direct testing of this hypothesis, and likely inform future research on the core mechanisms underlying endogenous control of prepotent impulses.

Acknowledgements

This work was supported by the European Science Foundation's EUROVETO project (09-ECRP-020).

REFERENCES

Achtziger, A., Alós-Ferrer, C., & Wagner, A. K. (2011). Social Preferences and Self-Control. Working Paper, University of Constance.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–277.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, *65*, 550–562.

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biological Psychiatry*, *69*, 55–68.

Aron, A. R., Robbins, T. W., & Poldrack, R. a. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *4*, 177-185.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*, 170–177.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science, 16,* 351-355.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nature Neuroscience*, *8*, 1458–1463.

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience*, *19*, 5473–5481.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*, 539–546.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *Journal of Neuroscience*, *27*, 9141–9145.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist*, *14*, 319–325.

Brass, M., Lynn, M. T., Demanet, J., & Rigoni, D. (2013). Imaging volition: what the brain can tell us about the will. *Experimental Brain Research*, *229*, 301–312.

Brody, A. L., Mandelkern, M. A., Olmstead, R. E., Jou, J., Tiongson, E., Allen, V., Scheibal, D., London, E. D., Monterosso, J. R., Stephen, T. T., Korb, A., Gan, J. J. & Cohen, M. S. (2007). Neural substrates of resisting craving during cigarette cue exposure. *Biological Psychiatry*, *62*, 642–651.

Campbell-Meiklejohn, D. K., Woolrich, M. W., Passingham, R. E., & Rogers, R. D. (2008). Knowing when to stop: the brain mechanisms of chasing losses. *Biological Psychiatry*, *63*, 293–300.

Chikazoe, J., Jimura, K., Hirose, S., Yamashita, K., Miyashita, Y., & Konishi, S. (2009). Preparation to inhibit a response complements response inhibition during performance of a stop-signal task. *Journal of Neuroscience*, *29*, 15870–15877.

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activation of right inferior frontal gyrus during response inhibition across response modalities. *Journal of Cognitive Neuroscience*, *19*, 69–80.

Craig, A. D. (2009). How do you feel-now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*, 59-70.

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*, 373–385.

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2003). The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response. *NeuroImage*, *20*, 404–412.

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*.

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*, 764–776.

Deiber, M. P., Honda, M., Ibañez, V., Sadato, N., & Hallett, M. (1999). Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: effect of movement type and rate. *Journal of neurophysiology, 81*, 3065–3077.

Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: a longitudinal study. *Neuropsychologia, 29*, 1113-1127.

Demanet, J., De Baene, W., Arrington, C. M., & Brass, M. (2013). Biasing free choices: The role of the rostral cingulate zone in intentional control. *NeuroImage*, *72*, 207–213.

Filevich, E., Kühn, S., & Haggard, P. (2012). Intentional inhibition in human action: the power of “no”. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1107–1118.

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience*, *33*, 299–324.

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*, 388–98.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*, 17538–17542.

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience*, *11*, 3656–3666.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*, 459–473.

Fujita, K. (2011). On Conceptualizing Self-Control as More Than the Effortful Inhibition of Impulses. *Personality and Social Psychology Review, 15*, 352-366.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind.” *Trends in Cognitive Sciences, 7*, 77-83.

Gibbons, F. X., Gerrard, M., Lando, H. A., & McGovern, P. G. (1991). Social comparison and smoking cessation: The role of the “typical smoker.” *Journal of Experimental Social Psychology, 3*, 239-258.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *Journal of Cognitive Neuroscience*, *18*, 932–948.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2* 271-299

Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition and Emotion*, *13*, 551–573.

Hanlon, C. A., Hartwell, K. J., Canterberry, M., Li, X., Owens, M., LeMatty, T., Prisciandaro, J. J., Borckardt, J., Brady, K. T. & George, M. S. (2013). Reduction of cue-induced craving through realtime neurofeedback in nicotine users: The role of region of interest selection and multiple visits. *Psychiatry Research: Neuroimaging*, *213*, 79–81.

Hartwell, K. J., Johnson, K. A., Li, X., Myrick, H., Lematty, T., George, S., & Brady, K. T. (2012). Neural correlates of craving and resisting craving for tobacco in nicotine dependent smokers. *Addiction and Biology, 16*, 654–666.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*, 122–128.

Jahfari, S., Stinear, C. M., Claffey, M., Verbruggen, F., & Aron, A. R. (2010). Responding with restraint: what are the neurocognitive mechanisms? *Journal of Cognitive Neuroscience*, *22*, 1479–1492.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science, 336*, 95-98.

Krieghoff, V., Brass, M., Prinz, W., & Waszak, F. (2009). Dissociating what and when of intentional actions. *Frontiers in Human Neuroscience*, *3*.

Krieghoff, V., Waszak, F., Prinz, W., & Brass, M. (2011). Neural and behavioral correlates of intentional actions. *Neuropsychologia*, *49*, 767–776.

Kritikos, A., Breen, N., & Mattingley, J. B. (2005). Anarchic hand syndrome: bimanual coordination and sensitivity to irrelevant information in unimanual reaches. *Brain Research. Cognitive Brain Research*, *24*, 634–647.

Kühn, S., & Brass, M. (2009). When doing nothing is an option: the neural correlates of deciding whether to act or not. *NeuroImage*, *46*, 1187–1193.

Kühn, S., Gallinat, J., & Brass, M. (2011). "Keep calm and carry on": structural correlates of expressive suppression of emotions. *PloS One*, 2*6*.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: how the "veto-area" exerts control. *Human Brain Mapping*, *30*, 2834–2843.

Kühn, S., Haggard, P., & Brass, M. (2013). Differences between endogenous and exogenous emotion inhibition in the human brain. *Brain Structure & Function*.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science*, *303*, 1208–1210.

Lau, H., Rogers, R. D., & Passingham, R. E. (2006). Dissociating response selection and conflict in the medial frontal surface. *NeuroImage*, *29*, 446–451.

Lévesque, J., Eugène, F., Joanette, Y., Paquette, V., Mensour, B., Beaudoin, G., Leroux, J. M., Bourgouin, P., & Beauregard, M. (2003). Neural circuitry underlying voluntary suppression of sadness. *Biological Psychiatry*, *53*, 502–510.

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*, *106*, 623–642.

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 259–289.

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*, 295-327.

Lynn, M. T., Van Dessel, P., & Brass, M. (2013). The influence of high-level beliefs on self-regulatory engagement: evidence from thermal pain stimulation. *Frontiers in Psychology*, 4.

Lynn, M. T., Demanet, J., Krebs, R. M., Van Dessel, P. & Brass, M. (in revision). Voluntary inhibition of pin avoidance behavior: An fMRI study.

Macrae, C. N., Christian, B. M., Golubickis, M., Karanasiou, M., Troksiarova, L., McNamara, D. L. & Miles, L. (in press). When do i wear me out? Mental simulation and the diminution of self-control. *Journal of Experimental Psychology: General.*

Mauss, I. B., Bunge, S. A., & Gross, J. J. (2007). Automatic emotion reguation. *Social and Personality Psychology Compass, 1*, 146-167.

Mischel, W. (1974). Processes in Delay of Gratification. *Advances in Experimental Social Psychology*, *7*, 249–292.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., Kross, E., Teslovich, T., Wilson, N. L., Zayas, V. & Shoda, Y. (2011). "Willpower" over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*, 252–256.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*, 655–663.

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*, 1354–61.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*, 856–69.

Neubert, F., Mars, R. B., Buch, E. R., Olivier, E., & Rushworth, M. F. S. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proceedings of the National Academy of Sciences, 107*, 13240-13245.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D. E., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, *23*, 483–499.

Pacherie, E. (2007). The anarchic hand syndrome and utilization behavior: a window onto agentive self-awareness. *Functional Neurology*, *22*, 211–217.

Passingham, R. E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14*, 16–21.

Passingham, R. E., Toni, I., & Rushworth, M. F. (2000). Specialisation within the prefrontal cortex: the ventral prefrontal cortex and associative learning. *Experimental Brain Research*, *133*, 103–113.

Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology*, *11*, 663–72.

Reuter, J., Raedler, T., Rose, M., Hand, I., Gläscher, J., & Büchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. *Nature Neuroscience*, *8*, 147–148.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.

Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition*, *21*, 1482–1490.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, *8*, 410–417.

Schel, M. A., Kühn, S., Brass, M., Haggard, P., Ridderinkhof, K. R., & Crone, E. a. (2014). Neural correlates of intentional and stimulus-driven inhibition: a comparison. *Frontiers in Human Neuroscience*, *8.*

Severens, E., Kühn, S., Hartsuiker, R. J., & Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Social Cognitive and Affective Neuroscience*, *7*, 431–435.

Sumner, P., Nachev, P., Morris, P., Peters, A. M., Jackson, S. R., Kennard, C., & Husain, M. (2007). Human Medial Frontal Cortex Mediates Unconscious Inhibition of Voluntary Action. *Neuron*, *54*, 697–711.

Toni, I., Rushworth, M. F., & Passingham, R. E. (2001). Neural correlates of visuomotor associations. Spatial rules compared with arbitrary rules. *Experimental Brain Research*, *141*(3), 359–369.

Van en Bos, W., McClure, S. M., Harris, L. T., Fiske, S. T. & Cohen, J. D. (2007). Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex. *Cognitive, Affective, and Behavioral Neuroscience, 7*, 337-346.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*, 829–858.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418–424.

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and Biobehavioral Reviews*, *33*, 647–61.

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*, 619–626.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience*, *7*, 1259–65.

Whitfield-Gabrieli, S., Moran, J. M., Nieto-Castañón, A., Triantafyllou, C., Saxe, R., & Gabrieli, J. D. E. (2011). Associations and dissociations between default and self-reference networks in the human brain. *NeuroImage*, *55*, 225–232.

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 11382–11387.

Zandbelt, B. B., Bloemendaal, M., Neggers, S. F. W., Kahn, R. S., & Vink, M. (2013). Expectations and violations: delineating the neural network of proactive inhibitory control. *Human Brain Mapping*, *34*, 2015–2024.

CHAPTER 3

Neural underpinnings of intentional inhibition reflect individual differences in self-regulation[1]

Behavioral inhibition is thought to be pivotal for successful self-regulation. Although the past decade has seen extensive research on the neural basis of inhibitory control, little is known about how findings from the laboratory relate to self-regulatory behaviors in real life. This is because most experimental paradigms either provide external information that fully determines the decision to act or inhibit, or employ rather arbitrary decision scenarios absent subjective behavioral urges, limiting their ecological validity. We recently introduced a novel paradigm that sought to overcome these limitations by requiring participants to intentionally inhibit the highly prepotent pain avoidance response. Having characterized the neural mechanisms involved in the inhibition of pain avoidance behavior, we now examine to what extent inhibition-related brain activity in our paradigm is related to individual differences in self-regulatory abilities. To this end, we examined self-regulatory abilities both at the trait level (by means of self-report) and at the state level (by means of an ego depletion manipulation). Replicating our previous findings, we found that inhibition of the pain avoidance response recruits a widespread network of inferior frontal, medial frontal, insular and posterior brain regions. Importantly, activity in core components of this network was modulated by trait levels of self-control. That is, participants with high self-regulatory abilities recruited these areas more strongly for inhibition. By contrast, no effect of our state manipulation was evident, potentially indicating problems with transference of ego depletion manipulations to the scanner environment. Implications of our findings for theories of behavioral inhibition and self-control are discussed.

[1] In collaboration with Paul S. Muhle-Karbe, Ruth M. Krebs, Jelle Demanet, and Marcel Brass.

INTRODUCTION

Behavioral inhibition is a prominent means by which successful self-regulation can be expressed (Baumeister, 2014). Not only are such omissions a daily struggle for every human being, but a number of neuropsychiatric conditions are also characterized by a lack of inhibitory control over unwanted yet strong behavioral impulses (e.g., Tourette's syndrome, Kwak, Dat Vuong, & Jankovic, 2003; anarchic hand syndrome, Marchetti & Della Sala, 1998; and addiction, Smith, Mattick, Jamadar, & Iredale, 2014). Thus, revealing the neural basis of inhibitory control over impulsive behavior is key both for understanding the mechanisms of adaptive self-regulation, and for advancing clinical interventions. That being said, behavioral inhibition is a concept that has proven notoriously challenging to study.

In cognitive neuroscience, inhibitory control has classically been related to the integrity of the right inferior frontal gyrus (rIFG, see Aron, Robbins, & Poldrack, 2014 for a recent review). A host of evidence indicates that the rIFG plays a pivotal role in the suppression of prepared actions in response to the occurrence of external stop or no-go signals (e.g., Chambers, Garavan, & Bellgrove, 2009; Forstmann et al., 2008; Wiecki & Frank, 2013), presumably in conjunction with the pre-supplementary motor area (preSMA) and the subthalamic nucleus (Aron & Poldrack, 2006). However, the transfer of these findings to self-regulation in everyday life is questionable, as the decision to inhibit is fully determined by external stimuli in these paradigms (henceforth referred to as *externally guided inhibition*). In realistic scenarios, on the other hand, the decision to withhold an impulsive act must typically originate endogenously. In an attempt to overcome this limitation, the concept of *intentional inhibition* has recently been introduced to the literature (Brass & Haggard, 2007). Intentional inhibition refers to the cancellation of behavior on the basis of free choice and thus strongly emphasizes the intentionality of inhibitory control. In the first studies to investigate this capacity, participants were asked to chose between either performing a simple action (e.g., pressing a button) or preparing the same movement but withholding it at the last possible moment (e.g., Kühn, Haggard, & Brass, 2009). Under these conditions, inhibition-related brain activity was found in the dorsal frontal median cortex (dFMC), but not in the rIFG, indicating that intentional inhibition may rely on neural mechanisms distinct from those of externally guided inhibition. However, the transfer of these

findings to self-regulation is still a difficult gap to bridge. Although decisions to suppress behavior were generated endogenously, they were also rather arbitrary (e.g., pressing a button or not), leaving participants with little prior motivation to act. This is in strong contrast with self-regulation in real life, which is characterized by the presence of a strong behavioral urge that is in conflict with individual goals (see Lynn, Muhle-Karbe, & Brass, 2014; Lynn, Van Dessel, & Brass, 2013).

We recently introduced a novel paradigm that aimed at combining these two critical aspects of behavioral inhibition in self-regulation: intentionality and the presence of behavioral urges (Lynn et al., 2013; Lynn et al., under revision). To this end, we investigated the suppression of the pain avoidance response (PAR). Thermal pain stimulation was applied to alternating inner wrists and participants could terminate the stimulation via a button press with the non-stimulated hand. On some trials, an external cue informed participants about the required response (externally guided trials), whereas on other trials participants could freely chose whether or not to persist terminating the stimulation (choice trials). The advantage of this paradigm is that Inhibition requires the suppression of a strong behavioral urge that is present on each trial and in each participant. Intriguingly, under these motivationally salient conditions, inhibition-related activation was found in the rIFG *and* in the dFMC, regardless of whether the decision to inhibit the PAR was externally guided or intentionally chosen. This finding strongly indicates that the motivational context is crucial when investigating behavioral inhibition. Moreover, it emphasizes that the brain regions implicated in externally guided and intentional inhibition, previously thought to be involved in largely independent functions, might operate in concert when genuine self-control needs to be exerted.

In the present study, we wanted to elaborate on this notion in more depth by relating inhibition-related activity in our paradigm to inter-individual differences in self-regulation. To this end, we employed two experimental approaches examining self-control both at the state level and the trait level. First, to examine self-control at the state level, we employed an ego depletion manipulation. Ego depletion is a concept from social psychology and refers to the process by which depletion of a limited self-regulatory resource during one effortful task leads to impediment of self-control in a subsequent, unrelated task (Baumeister, Bratslavsky, Muraven, & Tice, 1998). Hence,

prior to scanning, half of our participants we assigned to an effortful depletion condition, whereas the other half completed a less demanding control condition (see methods section). Second, to examine self-control at the trait level, we administered the Self-Regulation Questionnaire (Brown, Miller, & Lawendowski, 1999) to all participants one week before scanning. This questionnaire encompasses seven processes subserving successful self-regulatory behaviors in everyday life as defined by Miller & Brown (1991). We hypothesized that higher degrees of self-regulatory ability, both at the state level and at the trait level, would be reflected in stronger activation of self-control related brain regions when inhibiting the PAR.

METHODS AND MATERIALS

Participants

Forty Dutch-speaking participants (15 males) took part in the study (mean age = 22.38 years, SD = 4.07); each reported as healthy and with no history of neurological, pain, or circulatory disorders. All participants gave written informed consent, and the study was approved by the Medical Ethical Review Board of the Ghent University hospital, in accordance with the declaration of Helsinki. Participants were right-handed, as assessed by the Edinburgh Inventory (Oldfield, 1971), and were compensated thirty-five euros for their participation. Two participants had to be removed from all analyses due to severe head movement artefacts (>5mm).

Experimental procedure

Methods were similar to those of Lynn et al. (under revision), with two important exceptions. First, because no substantial differences were found between intentional and externally-guided inhibition in the prior study, we opted to use only free choice trials here to increase the statistical power of our design. This also raises the interesting question of whether similar results would be found in an unconstrained context of exclusively free choice trials, or whether a mixture of endogenous and exogenous decisions are necessary to elicit the activation patterns found in Lynn et al. (under revision). Second, the data of the present study were collected as part of a larger experimental endeavor, in which the action-outcome contingency was manipulated in one of the two blocks (counterbalanced). In this second 'non-

contingent' block, the outcome of a response was determined randomly, such that the pain stimulation would occasionally persist following a button press, or conversely, terminate early despite the lack of a response, thus eliminating the expected response contingency[2]. This paper is based only on the data of the contingent block because here we are interested in relating the previously established patterns of inhibition-related activity to self-control. Data of the non-contingent block will be presented in an independent report.

Pain tolerance threshold determination. Pain was induced via a thermode connected to a Medoc PATHWAY device (MEDOC, Haifa, Israel), an apparatus designed for the induction of thermal pain using cold or hot stimulation. The temperature at which participants felt a sufficient amount of pain was determined during a pre-test session taking place one week prior to scanning. Participants were exposed to 26 trials in which the thermal sensation gradually increased over five seconds from 32°C to a randomized destination temperature between 45 and 50°C (in increments of .25 degrees), a slope comparable to that of the experimental trials. After each trial, the thermode returned instantly to baseline temperature, and participants were asked to rate their perceived pain on a scale from zero to eight, with zero being no pain and eight being the worst possible pain. The destination temperature employed in the main experiment was computed for each participant as the highest temperature at which they rated their pain as a six[3]. This method was revealed during piloting to yield more accurate tolerance threshold measurements than merely requiring participants to indicate the maximum heat they could withstand when exposed to a steadily increasing temperature. Importantly, participants were free to press a button at any point during the threshold determination in order to terminate the trial.

Depletion manipulation. The day of scanning, participants were asked to rate their affect and tiredness upon arrival. Valence of affect was measured with the Self-Assessment Manikin (SAM; Lang, 1980), while tiredness was rated on a scale of 0 to 10, with 0 indicating "not tired at all" and 10 indicating "very tired." Participants then completed either a depletion task, designed to exhaust limited self-regulatory

[2] This was done to examine the effect of outcome anticipation on the neural basis of inhibitory self-control.

[3] This temperature is henceforth referred to as the individual participant's tolerance threshold.

resources, or a control task, described below. Following the task, they rated how difficult the task was, and how much effort it took to complete the task (both on a scale of 0 to 10, with 0 indicating "not difficult/no effort at all" and 10 indicating "very difficult/very much effort"). The efficacy of the depletion manipulation was then evaluated prior to scanning with an anagram task (see Baumeister et al., 1998, experiment 3, for a similar task). Participants were asked to solve ten six-letter anagrams, two of which were unsolvable nonsense words. Briefer persistence at an unsolvable task and a reduced number of anagrams solved are taken as indicators of depletion. Again, following completion of the anagram task, participants rated the task's difficulty and effort.

When selecting a depletion task, our primary concern was finding a way to exhaust participants' self-regulatory resources without using what would strictly be classified as an inhibition task. Traditionally, ego depletion studies most often employ tasks such as suppressing thoughts and emotions (see Hagger, Wood, Stiff, & Chatzisarantis, 2010 for a review). However, these tasks would introduce a confound to our design when examining inhibition-related brain activation as the dependent measure. Based on several pilot studies[4], we therefore chose a different depletion method: participants solved math problems, consisting of addition and multiplication of one- to three-digit numbers. They were instructed to solve as many math problems as possible within a twenty-minute time period. Those in the depletion group were given several additional instructions: they should complete the problems in order (no skipping to the easiest problems), they should only write solutions on the paper, not their work (this should be maintained in working memory), and finally they were given headsets to wear through which they heard a computer-generated voice randomly naming single-digit numbers. This was meant to increase the cognitive load for ego depleted but not control participants.

***Task and stimuli*.** Participants received thermal pain stimulation during each trial, applied via a thermode to alternating inner wrists. Each trial began with the

[3] We also piloted a depletion manipulation with a variant of the so-called "e-hunting" task where participants are asked to cross out all instances of the letter 'e' in a printed narrative for a longer time period (see Debey, Verschuere, & Crombez, 2012). However, as the manipulation with the math problems was rated as more difficult and produced stronger effects on the anagram task, we opted to use the math task as the depletion manipulation in the present study.

presentation of a fixation cross for five seconds, during which time the temperature of the thermode began to gradually increase from a baseline of 32°C to the participant's individually determined tolerance threshold. Subsequently, a cue appeared in place of the fixation cross, and persisted for the remainder of the pain stimulation. This cue indicated that participants should select one of the two possible response options, and either press the response button immediately, or persist and suppress the urge to withdraw from the pain sensation ('choice action' and 'choice inhibition,' respectively). Participants were requested to make their response decision at the moment of cue onset and to choose both options approximately equally often over the course of the experiment, but not to use any particular strategies (e.g., simple response alternations) or decide in advance of the presentation of the cue. Adherence to these instructions was subsequently assessed by calculating each participant's Random Number Generation 2 (RNG2; an index optimized for two-choice response sequences) score using the program RgCalc (Towse & Neil, 1998). RNG2 scores range from 0 to 1, with 1 indicating complete sequence predictability. The temperature remained at tolerance threshold for the next three seconds, or until the participant pressed the button to terminate both the pain stimulation and the trial. Participants responded with the index or middle finger of the arm not being stimulated (thereby providing a response time for action trials). This was followed by a six second rest period. Afterwards, prompts were presented for two seconds each, asking participants for verbal ratings (collected via a microphone inside the scanner bore) of their perceived pain and their subjective 'urge to terminate the trial by pressing the button' (both on a scale of 0 – 8). Participants were then cued to alternate the arm placed atop the thermode, and were given 10 seconds in which to accomplish this task with the assistance of an experimenter who stood near the scanner bore. An experimenter, present in the scanner room throughout the session, also placed a small sandbag over the to-be stimulated wrist in order to lend weight and prevent the participant from inadvertently withdrawing from the pain source rather than button pressing. Each trial ended with an additional six-second rest period. A schematic overview of the trial structure is presented in Figure 1. Each participant had to perform forty trials per block. Importantly, participants were free to press a button to immediately terminate the thermal sensation at any point during the experiment.

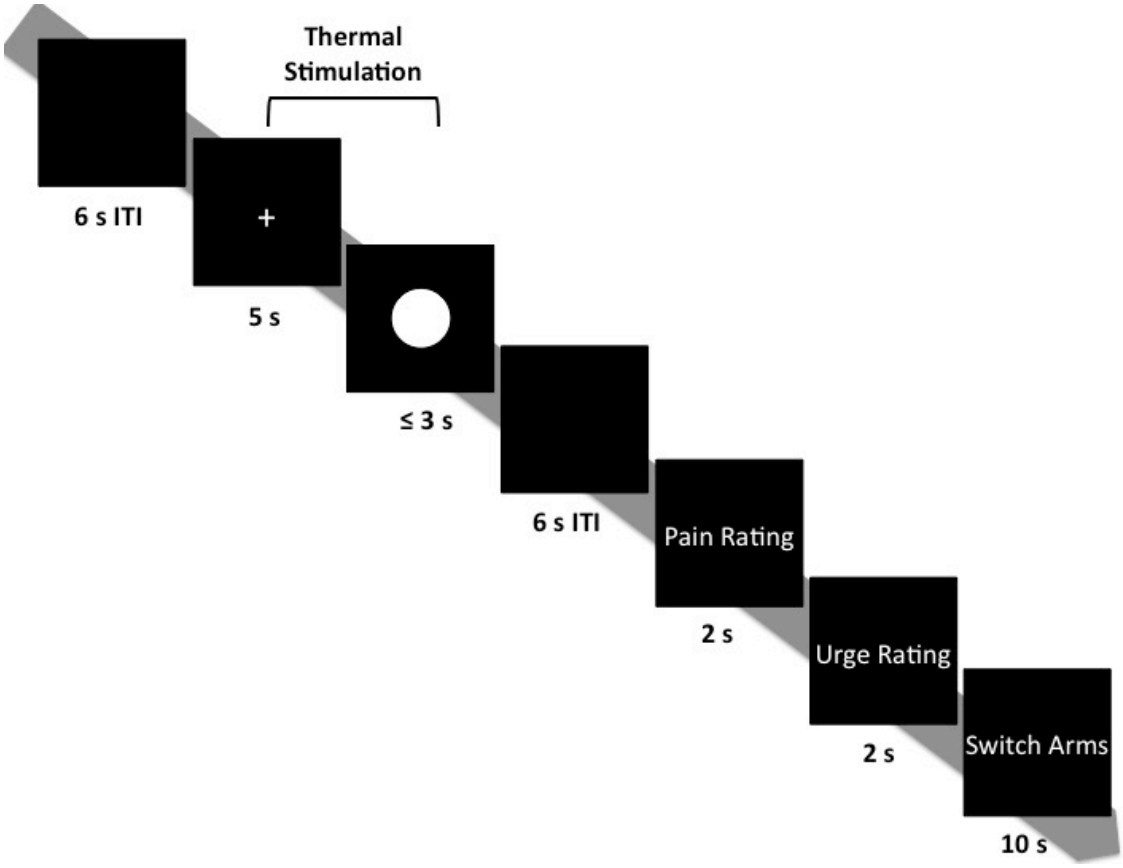

Figure 1. ***Schematic of a single trial.***

Post-experimental debriefing and questionnaires. Following the scanning session, participants were asked to again rate their mood and tiredness. They also completed a funneled debriefing, which probed whether they used any particular strategies during the experiment, whether they made their decisions in advance of the cue, what proportion of inhibition they believed they performed, and their subjective impression of the sensations elicited by the thermal stimulation.

Behavioral data preparation

Trials in which participants did not place their wrist directly on the thermode were indicated with pain and urge ratings of zero and were excluded from behavioral and fMRI analysis. Likewise trials with reaction times above 2000 ms were removed, in order to use a response window comparable to that of Lynn et al. (in revision) and because we assume that such extremely lengthy delays before responding do not reflect accurate following of the instructions to decide spontaneously and immediately at the occurrence of the cue. Errors in the form of null ratings led to the exclusion of 18

trials (1520 overall), all but one classified as inhibition trials (number of excluded trials per participant: *M* = 0.47, *SD* = 1.25) from both behavioral and fMRI analysis. Errors in the form of delayed response times led to the exclusion of 65 action trials (number of excluded trials per participant: *M* = 1.71, *SD* = 2.80) from both behavioral and fMRI analysis. Pain and urge ratings were analyzed using repeated-measures ANOVAs with RESPONSE (action vs. inhibition) as a within-subjects factor and DEPLETION (control vs. depleted) as a between-subjects factor. Reaction times on action trials and the relative proportion of inhibition trials were analyzed as a function of DEPLETION via two-tailed independent-samples *t*-tests.

fMRI data acquisition and preprocessing

Data were acquired with a 3T Siemens Magnetom Trio MRI system (Siemens Medical Systems, Erlangen, Germany) using a 32-channel radiofrequency head coil. Subjects were positioned headfirst and supine in the magnet bore. First, 176 high-resolution anatomical images were acquired using a T1-weighted 3D MPRAGE sequence (TR = 2,250 ms, TE = 4.18 ms, TI = 900 ms, image matrix = 256 x 256, FOV = 256 mm, flip angle = 9°, and voxel size = 1 x 1 x 1 mm). Whole-brain functional images were then collected using a T2-weighted echo-planar imaging (EPI) sequence, sensitive to blood-oxygen-level dependent contrast (TR = 2,000 ms, TE = 35 ms, image matrix = 64 x 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 17%, voxel size 3.5 x 3.5 x 3.0 mm, and 30 axial slices). A varying number of images were acquired per run due to individual differences in choice behavior and reaction times.

All data were preprocessed and analyzed using Matlab and the SPM8 software (Wellcome Department of Cognitive Neurology, London, UK). To account for possible T1 relaxation effects, the first four scans of each EPI series were excluded from the analysis. First, a mean image for all scan volumes was created, to which individual volumes were spatially realigned using rigid body transformation. Thereafter, they were slice time corrected using the first slice as a reference. The structural image of each subject was coregistered with their mean functional image after which all functional images were normalized to the Montreal Neurological Institute (Montreal, Quebec, Canada) T1 template. The images were resampled into 3.5 x 3.5 x 3.5 mm voxels and spatially smoothed with a Gaussian kernel of 8 mm (full-width at half maximum). A

high-pass filter of 128 Hz was applied during fMRI data analysis.

Statistical analyses

The first-level statistical analyses were performed using a general linear model (GLM, Friston et al., 1995). In order to distinguish the pre-cue buildup of urge from the moment of decision and subsequent inhibition, we modeled the five-second pre-cue epoch and used the onset of the task cue as the main event of interest in the GLM. It is important to note that all trials were identical in terms of stimulation up to and including cue onset, and for an average of 740 ms afterwards (i.e., average time it took to respond in action trials). Four regressors were defined reflecting the experimental conditions and the pre-cue epochs for each experimental condition (i.e., 'action pre-cue,' 'inhibition pre-cue,' 'action cue onset,' 'inhibition cue onset'). Six regressors defining head movements were included to account for any residual movement-related effects. All regressors were convolved with a canonical hemodynamic response function (HRF). Contrast images were computed separately for each participant to compare parameter estimates of the relevant conditions.

These contrast images were then advanced to the second level, using a one-way ANOVA design as implemented in SPM8 with the factor RESPONSE (action vs. inhibition). We examined the main effect of inhibition to reveal brain areas that were more active on inhibition trials than on action trials. To control for false-positive rates, combined voxel activation intensity and cluster extent thresholds corrected for multiple comparisons were determined using 3dClustSim (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html). This widely used correction method is applied to statistical contrast images at the group level and estimates the probability of observing false positive (random fields of noise) clusters of a given size, as a function of a given voxelwise *p* value. The 3dClustSim program considers the size of the image (number of voxels), the voxelwise statistical values, and the spatial correlations over voxels (spatial smoothness) and runs a user-specified number of Monte Carlo simulations to generate an appropriate null-distribution. Here, ten thousand Monte Carlo simulations were run, taking into account the whole-brain search volume and the estimated smoothness of each axis (*x, y,* and *z*) of the respective group SPMs. Probability estimates of a random field of noise were generated,

producing a cluster of voxels of a given extent for a set of voxels passing a voxelwise *p* value threshold of 0.001. Given this voxelwise threshold, the simulations determined that cluster sizes of 20.0 – 23.5 voxels, depending on the specific contrast analysis, corresponded to a combined threshold of $p < 0.05$ (corrected). Whole-brain analyses were supplemented with region-of-interest (ROI) analyses. ROIs were generated using MARSBAR toolbox for use with SPM 8 (Brett, Anton, Valabregue, & Poline, 2002).

RESULTS

Behavioral results

Depletion manipulation. Participants in the depletion group rated the depletion manipulation itself (math problems) as more difficult, $t(36) = 3.290$, $p = .002$, and requiring more effort than did those in the control group, $t(36) = 2.158$, $p = .038$ (Math difficulty: Depleted $M = 7.353$, $SD = 1.295$; Control $M = 5.774$, $SD = 1.643$; Math effort: Depleted $M = 7.763$, $SD = 1.336$; Control $M = 6.690$, $SD = 1.708$). However, there were no significant differences between groups in terms of anagram difficulty and effort, pretest mood and tiredness, or posttest mood and tiredness, all *p*-values > .370. Computing difference scores for posttest – pretest mood and tiredness yielded no significant effect of tiredness, $p = .715$. There was a trend for a difference between groups in the mood difference score, such that ego depleted participants exhibited less reduction in mood over the course of the experiment, $t(36) = 1.952$, $p = .059$ (Depleted $M = .526$, $SD = 1.429$; Control $M = 1.368$, $SD = 1.223$). In terms of anagram performance, there were no significant differences between groups in terms of persistence at unsolvable anagrams (Depleted RT $M = 63.068$ s, $SD = 37.122$; Control RT $M = 67.099$ s, $SD = 23.388$), or in number of solvable anagrams completed (Depleted $M = 5.737$, $SD = 1.593$; Control $M = 6.053$, $SD = 1.268$), both *p*-values > .503. Overall, this pattern indicates that the manipulation was not effective at depleting participants' self-regulatory resources.

Proportion of choices. Participants chose to inhibit, on average, in 60.32% of trials (Range = 45.0% - 97.3%; $SD = 11.64$). Ego depleted and control participants did not differ in terms of their proportion of inhibition, $t(36) = .408$, $p = .686$.

Randomness of choice response sequences. Participants displayed a mean RNG2 index of 0.732 ($SD = .024$). Individual scores were compared to thirty-eight randomly

generated sets of two-choice response sequences (RNG2 *M* = 0.723, *SD* = .010) in an independent samples *t*-test. Participants' choice trial responses differed significantly from the randomly generated samples [*t*(51) = 2.020, *p* = .049; Levene's test indicated unequal variances, so degrees of freedom were adjusted from 74], in that their choices were slightly more predictable. However, these values are comparable to previous studies on intentional inhibition (e.g., Schel et al., 2014), and indicate that participants' choice behavior, albeit not entirely random, cannot be explained by simple alternation strategies. Ego depleted and control participants did not differ in their RNG2 scores, *t*(36) = .009, *p* = .993 (Depleted *M* = .732, *SD* = .029; Control *M* = .732, *SD* = .018).

Reaction times. Participants had a grand mean reaction time of 739.76 ms on action trials. Depleted and control participants did not differ significantly in their reaction times, *t*(36) = 1.081, p = .287 (Depleted *M* = 696.71, *SD* = 208.46; Control *M* = 782.80, *SD* = 277.49).

Subjective pain ratings. The main effect of RESPONSE (Action vs. Inhibition) was not significant, *F*(36) = .019, *p* = .892 (Action *M* = 4.780, *SD* = 1.155; Inhibition *M* = 4.820, *SD* = 1.016). However, there was a trend for an interaction between RESPONSE and DEPLETION (Depleted vs. Control), *F*(36) = 3.848, *p* = .058, such that depleted participants rated action trials as more painful than inhibit trials (Action *M* = 4.982, *SD* = 1.127; Inhibition *M* = 4.706, *SD* = 0.942), and control participants rated inhibit trials as more painful than action trials (Action *M* = 4.617, *SD* = 1.183; Inhibition *M* = 4.934, *SD* = 1.100). This might indicate that depleted participants were slightly more likely to make their choices according to pain at the moment of decision.

Subjective urge ratings. There was a significant main effect of RESPONSE, *F*(36) = 13.272, p = .001, such that participants rated their urge to press the response button in order to terminate the trial in accordance with their decisions (Action *M* = 4.610, *SD* = .280; Inhibition *M* = 3.865, *SD* = .207). The interaction between RESPONSE and DEPLETION was not significant, *p* = .478.

fMRI results

Whole-brain analyses. We first analyzed the main effect of RESPONSE to reveal the brain areas that exhibit stronger activation on inhibition trials than on action trials (see Figure 2). Activity was found bilaterally within the IFG and the adjacent insular

cortex. Additional clusters were located in the anterior cingulate cortex, preSMA, inferior parietal lobule (IPL), and right dorsolateral prefrontal cortex (dlPFC), as well as in the globus pallidus and thalamus. For a complete overview of significant activations, see Table 1.

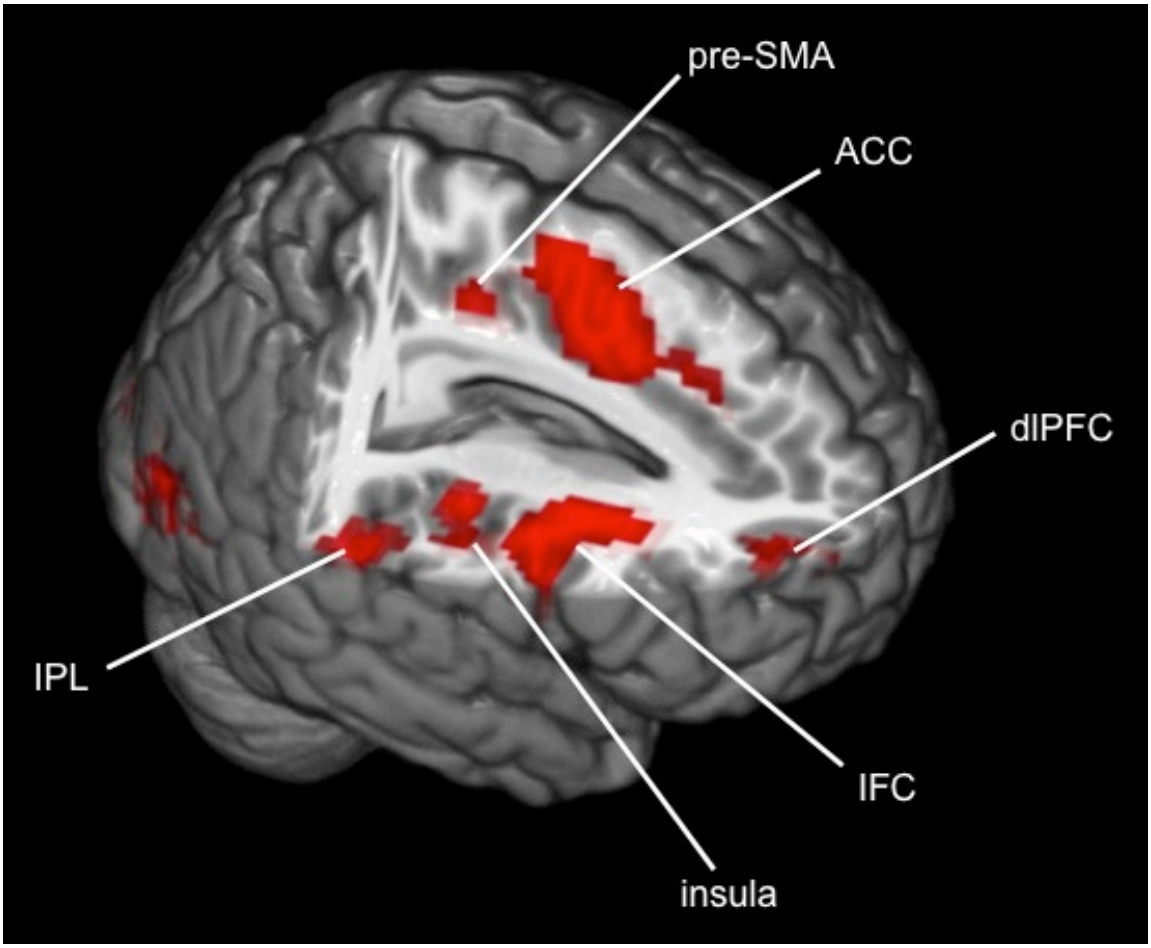

Figure 2. ***Activation map of the whole-brain contrast comparing inhibition versus action. Note the map was thresholded at p < .001 (uncorrected), with a minimum cluster size of 22 contiguous voxels (see Methods section for details).***

Table 1

Overview of activation clusters revealed by the whole-brain analyses. Montreal Neurological Institute (MNI) coordinates reflect the peak voxel of a given cluster.

Region	Voxels	MNI coordinates			*Tmax*
		x	*y*	*z*	
Main effect inhibition: inhibit > press					
Right anterior insula	1176	33	26	4	7.69
Inferior frontal gyrus		54	5	10	6.42
Posterior insula		36	8	13	6.19
Anterior cingulate cortex	605	12	2	49	6.29
Pre-supplementary motor area		12	-1	58	5.80
Cingulate gyrus		12	-25	43	5.04
Left inferior frontal gyrus	314	-54	2	10	6.01
Anterior insula		-33	17	10	5.01
Left inferior parietal lobule	110	-60	-25	25	5.06
Posterior insula		-36	-19	19	4.74
Right dlPFC	174	33	44	31	5.04
		24	56	10	3.58
Globus pallidus	35	18	-1	-5	4.55
Putamen		18	8	-5	3.99
Thalamus	51	6	-16	1	4.19
Red nucleus		6	-16	8	4.11
Right inferior occipital gyrus	431	45	-73	-5	8.05
Left inferior occipital gyrus	180	-45	-79	-5	5.44

Region-of-interest (ROI) analyses. Next, we conducted a series of ROI analyses in order to examine trait self-regulatory effects. We selected only ROIs that were observed in both the present study and that of Lynn et al. (under revision) for the whole-brain inhibition vs. action contrast, resulting in the following ROIs: the insula and IFG (each bilateral in the present study), right IPL, and preSMA. Spherical ROIs with a radius of 6mm were drawn for the smaller clusters (i.e., IFG and IPL), while ROIs of larger clusters had a radius of 10mm (i.e., insula and preSMA). ROIs were centered on the peak MNI coordinates reported in Table 1. We also created an ROI for the dFMC based on the coordinate reported by Kühn et al. (2009; radius = 6 mm, MNI peak -7 42 21), given the consistent role of this region in previous intentional inhibition studies. Each participant's SRQ score was then computed and used as a between-subjects factor, dividing participants into quartiles of low to high trait self-regulatory ability. For each ROI, a repeated-measures ANOVA was conducted, with percent signal change on inhibition and action trials as a within-subjects factor and the SRQ Quartile as a covariate. Significance thresholds were corrected for multiple comparisons using the Holm-Bonferroni method (Holm, 1979). The left IFG showed a main effect of Response, $F(1,34) = 36.046$, $p < .001$, and an interaction between Response and SRQ Quartile, $F(3,34) = 6.539$, $p = .001$, wherein participants with higher trait self-regulation showed greater inhibition-related activation than participants with low trait self-regulation (see Figure 3). The left anterior insula showed the same pattern of results, with a main effect of Response, $F(1,34) = 29.788$, $p < .001$, and an interaction between Response and SRQ Quartile, $F(3,34) = 6.876$, $p = .001$. The dFMC showed a significant interaction between Response and SRQ Quartile, $F(3,34) = 4.865$, $p = .006$, yet no main effect of Response, $p = .862$. As Figure 4 indicates, this pattern reflects that this region was strongly recruited for inhibition by participants with high self-regulatory ability, but the opposite seems to be true of participants with low self-regulatory ability. The main effect of Response was significant for the remaining four ROIs, all p-values < .001, reflecting increased activity on inhibition trials. The interaction terms for the right IFG and right anterior insula ROIs were marginally significant but did not survive correction for multiple comparison (uncorrected p-values = .017 and .024, respectively), and showed the same pattern of results as their left counterparts. The preSMA and IPL did not produce significant interactions between Response and SRQ Quartile, p-values > .230.

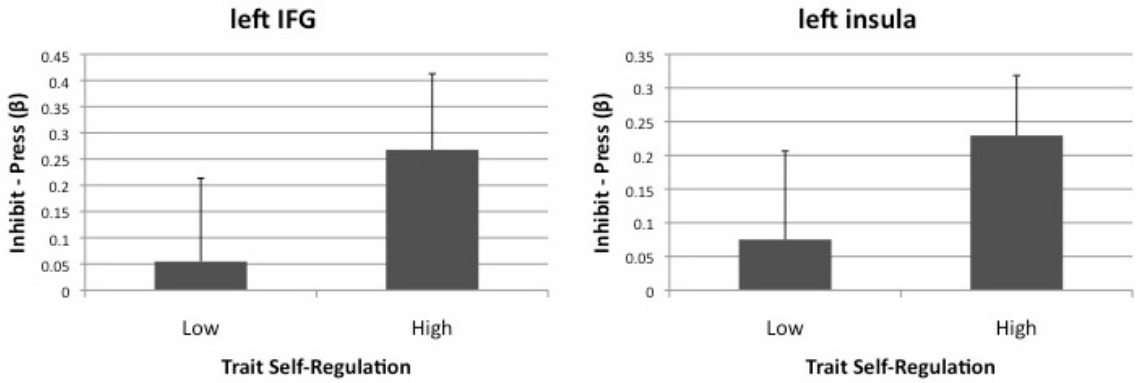

Figure 3. ***Left-lateralized results of the trait self-regulation ROI analysis. Based on the study of Lynn et al. (under revision), spherical ROIs were centered at the peak IFG and insula coordinates (-54 2 10, radius 6 mm and -33 17 10, radius 10 mm respectively). Values represent standard errors and a difference score of the mean percent signal change (beta values) derived from inhibit versus press trials.***

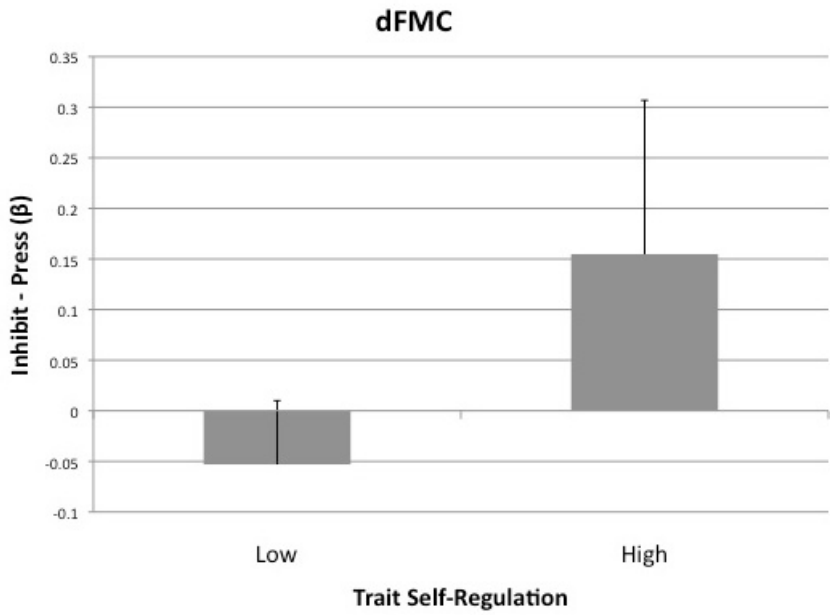

Figure 4. ***Frontomedian results of the trait self-regulation ROI analysis. A spherical ROI (radius = 6 mm) was centered at the peak dFMC coordinate (-7 42 21) resulting from the study of Kühn et al. (2009). Values represent standard errors and a difference score of the mean percent signal change (beta values) derived from inhibit versus press trials.***

In a second step, we examined state differences in self-control at the ROI level. Although the behavioral results indicated that the state self-control manipulation was not effective at altering participants' access to self-regulatory resources, we nevertheless inspected whether there were any group differences in the ROIs, by conducting two additional repeated-measures ANOVAs for each ROI with percent signal change on action and inhibition trials as a within-subjects factor. In the first ANOVA, DEPLETION (control vs. depleted) was entered as a between-subjects factor, and in the second ANOVA the difference score between unsolvable anagram RT and solvable anagram RT was used as a covariate (as a behavioral index of depletion). Neither analysis yielded a significant interaction for any of the seven ROIs (all *p*-values < .271).

DISCUSSION

In the present study, we investigated the neural basis of intentional inhibition in the context of pain avoidance behavior, and its relation to state and trait differences in self-regulatory ability. Largely replicating our previous results, we found a network to be activated during inhibition that comprised areas implicated in both externally guided and intentional inhibition. Importantly, we also observed that the degree to which these regions were recruited for inhibition was modulated by participants' trait level of self-regulation, highlighting a link between the neuroimaging literature on response inhibition and self-control in realistic scenarios. By contrast, no effects were found for our state manipulation, neither behaviorally nor at the brain level. Below we discuss the implications of our findings for the neuroscience of self-regulation and impulse control.

A network for inhibitory self-regulation

The present study replicated our previous finding that intentional inhibition of the PAR relies on a combination of brain areas that have been implicated in externally guided inhibition (e.g., the IFG, the preSMA, and the globus pallidus) and those implicated in intentional inhibition (e.g., dFMC, IPL, and anterior insula), although it needs to be mentioned that the dFMC could only be identified in a ROI analysis, and not in the whole-brain analysis. Nonetheless, this replication underscores that the original idea of two independent inhibition systems might only hold for specific experimental situations, in particular those where participants operate in "cold" response systems

with little motivational incentives. By contrast, in motivationally salient conditions that recruit "hot" response systems, successful stopping appears to require the interaction of both networks (see Lynn et al., under revision). This does, however, not imply that these areas serve equivalent functions in inhibitory control. We have previously proposed that the rIFG - in conjunction with the preSMA and subcortical areas such as the subthalamic nucleas and the globus pallidus - is critical primarily for the implementation of behavioral inhibition, i.e., the actual suppression of prepotent motor output (Lynn et al., 2014). The dFMC, on the other hand, appears to be involved in more abstract decisional aspects of inhibition, though it is still an ongoing debate as to whether this decision reflects a "veto" signal for the cancellation of motor output (see Filevich, Kühn, & Haggard, 2012), or instead higher-level disengagement from urges and intentions (see Lynn et al., 2014). Either way, we believe that when faced with the demand to suppress a very strong behavioral urge, such as the urge to avoid pain, both decisional and implementation aspects presumably need to be reinforced continuously to enable compliance with the instructions.

Linking brain activity with real-life self-regulation

Importantly, beyond strengthening the notion that voluntary inhibition of the PAR relies on joint activation of brain areas implicated in externally guided and intentional inhibition, the present study provides novel evidence that activation in these areas is related to self-regulatory abilities in real life. Participants with higher self-reported self-regulatory abilities exhibited stronger inhibition-related activation in the core components of the inhibition network (i.e., IFG, insula, and dFMC). This finding constitutes an important step in the endeavor to transfer findings from inhibition tasks in the lab to self-regulation in real life (see also O'Connor, Upton, Moore, & Hester, 2014; Wessel & Aron, 2014 for motivationally-salient investigations of self-control). Importantly, a host of social psychological research emphasizes that trait levels of self-control predict many important behavioral outcomes such as the extent to which impulses are indulged (Friese & Hofmann, 2009) and how well individuals perform on behavioral self-control tasks (Schmeichel & Zell, 2007). In general, our data suggest that a stronger recruitment of the inhibition network might underlie increased trait self-regulatory ability, but more research will be necessary to substantiate this view. For

instance, this could be achieved by examining patient groups with impulsive disorders, or in longitudinal studies that trace the development of self-regulatory abilities across childhood and adolescence.

Another interesting aspect of our data pertains to the fact that trait differences in self-regulation were more clearly reflected in activity of the left IFG than in activity of the right IFG. Although the left IFG has occasionally been implicated in response inhibition as well (e.g., Swick, Ashley, & Turken, 2008), the right IFG is generally considered more critical for this capacity (see Aron et al., 2014). The left IFG is typically assumed to be involved in more general cognitive control functions such as the representation of task rules (e.g., Hartstra, Kühn, Verguts, & Brass, 2011; Mecklinger, von Cramon, Springer, & Matthes-von Cramon, 1999). From this perspective, our finding could imply that behavioral stopping is not the sole distinctive feature of self-control, and that more general control mechanisms also contribute to successful self-regulation.

No evidence for state-effects of self-control

In contrast with the link between inhibition-related brain activity and trait measures of self-regulation, our state manipulation did not appear to impact the results. This null effect likely reflects that, despite extensive piloting, the depletion task was ineffective at reducing self-regulatory resources in the present experimental context and illustrates the challenge of applying social psychological concepts to neuroscience research. A number of reasons come to mind as to why the manipulation did not impact participants' self-control, neither behaviorally nor at the brain level: First, the stressful scanner environment may have led to a general increase in participants' arousal levels, making them less susceptible to subtle depletion effects than in behavioral sessions in a non-scanning environment. Previous studies have shown that the presentation of MR noise during task performance can lead to an increase of cognitive control and thereby diminish or even eliminate established experimental effects (Hommel, Fischer, Colzato, van den Wildenberg, & Cellini, 2012). Second, it is also possible that depletion effects were too short-lived to be measured in our setup given both the time lag between the completion of the math problems and the beginning of the inhibition task (following the collection of anatomical image), and

the overall duration of the scanning session (see Hagger et al., 2010, and Tyler & Burns, 2008 for evidence regarding the duration of ego depletion effects). Third, participants might furthermore have been thinking ahead and regulated their resources in advance of the manipulation, having full anticipatory knowledge of what would be required of them in the second part of the experiment. Finally, the choice of our depletion task, intended to avoid inhibitory demands, might have rendered this manipulation to be less powerful than traditional ego depletion tasks such as thought or emotion suppression. If correct, this would indicate that the transfer of ego depletion to other tasks is much narrower than previously assumed, and that depletion might depend on common cognitive functions being required in both the depletion task and the test task. Future research could test this idea by systematically varying the degree of functional similarity between both tasks. Overall, it is worth noting that the concept of ego depletion has recently been called into question. Carter & Mccullough (2014) suggest that the ego depletion effect is rather weak if not absent, while Xu et al. (2014) could not replicate the effect in four separate studies using the strongest of depletion protocols. Beedie & Lane (2011) negate prior assumptions about the role of glucose as willpower's physiological substrate. Finally, abundant evidence indicates that the ego depletion effect seems to rely on mediating factors such as an individual's beliefs about the veracity of the limited resource model and about their own level of depletion (Clarkson, Hirt, Jia, & Alexander, 2010; Job, Dweck, & Walton, 2010).

Future perspectives

Future studies that attempt to apply ego depletion to neuroimaging should take the above-mentioned factors into account. Those studies addressing impulse control in particular should attempt to ensure that experimental paradigms induce a strong action tendency in each subject, allow for the endogenous generation of the decision to inhibit, and establish a link between performance in the lab and self-regulatory behaviors outside the lab. Preferably, future studies would employ experimental contexts other than pain avoidance that fulfill these criteria. This could help to reveal to what extent our findings are driven by the presence of pain and to delineate the domain-specificity of the neural basis of self-control.

Conclusion

In the present study, we examined the neural underpinnings of the endogenous inhibition of pain avoidance behavior and its relation to self-regulatory abilities at both state and trait levels. Our results show inhibition-related activity in a previously established network, which was associated with participants' trait level of self-regulation. By contrast, no effects were found for the state manipulation of self-regulatory abilities. Together, these findings provide evidence that brain activity during response inhibition reflects individual differences in realistic self-regulatory behaviors and thus opens the door to a more ecologically valid investigation of the neural basis of inhibitory self-control.

Acknowledgements

This work was supported by the European Science Foundation's EUROVETO project (09-ECRP-020).

REFERENCES

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience*, *26*(9), 2424–2433.

Aron, A. R., Robbins, T. W., & Poldrack, R. a. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–85.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource?, *74*(5), 1252–1265.

Baumeister, R. F. (2014). Self-regulation, Ego Depletion, and Inhibition. *Neuropsychologia*, 1–7.

Beedie, C. J. & Lane, A. M. (2011). The role of glucose in self-control: Another look at the evidence and an alternative conceptualization. *Personality and Social Psychology Review.*

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(34), 9141–5.

Brett, M., Anton, J.-L. L., Valabregue, R., & Poline, J.-B. B. (2002). Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. In *NeuroImage* (Vol. 16, p. abstract 497).

Brown, J. M., Miller, W. R., & Lawendowski, L. A. (1999). The Self-Regulation Questionnaire (SRQ). In *Innovations in clinical practice: A source book (Vol. 17)* (pp. 281–289).

Carter, E. C. & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology,* 5:823, 1 – 11.

Chambers, C. D., Garavan, H., & Bellgrove, M. a. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews*, *33*(5), 631–46.

Clarkson, J. J., Hirt, E. R., Jia, L., & Alexander, M. B. (2010). When perception is more than reality: the effects of perceived versus actual resource depletion on self-regulatory behavior. *Journal of Personality and Social Psychology, 98*(1), 29 – 46.

Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: an experimental investigation using ego depletion and goal neglect. *Acta Psychologica*, *140*(2), 133–41.

Filevich, E., Kühn, S., & Haggard, P. (2012). Intentional inhibition in human action: the power of "no". *Neuroscience and Biobehavioral Reviews*, *36*(4), 1107–18.

Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2008). Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: a model-based approach. *The Journal of Neuroscience*, *28*, 9790–9796.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.

Friese, M. & Hoffman, W. (2009). Control me or I will control you: Impulses, trait self-control, and the guidance of behavior. *Journal of Research in Personality, 43*(5), 795 – 805.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychological Bulletin*, *136*(4), 495–525.

Hartstra, E., Kühn, S., Verguts, T., Brass, M. (2011). The implementation of verbal instructions: An fMRI study. *Human Brain Mapping, 32*(11), 1811 – 24.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Hommel, B., Fischer, R., Colzato, L. S., van den Wildenberg, W. P. M., & Cellini, C. (2012). The effect of fMRI (noise) on cognitive control. *Journal of Experimental Psychology: Human Perception and Performance*.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion – Is it all in your head?: Implicit theories about willpower affect self-regulation. *Psychological Science, 21*(11), 1686 – 93.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: how the "veto-area" exerts control. *Human Brain Mapping*, *30*(9), 2834–43.

Kwak, C., Dat Vuong, K., & Jankovic, J. (2003). Premonitory sensory phenomenon in Tourette's syndrome. *Movement Disorders*, *18*(12), 1530 – 1533.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119 – 137). Norwood, NJ: Ablex.

Lynn, M.T., Demanet, J., Krebs, R. M., Van Dessel, P., & Brass, M. (under revision). Voluntary inhibition of pain avoidance behavior: An fMRI study. *Brain Structure and Function.*

Lynn, M. T., Muhle-Karbe, P. S., & Brass, M. (2014). Controlling the self: The role of the dorsal frontomedian cortex in intentional inhibition. *Neuropsychologia*, 1–8.

Lynn, M. T., Van Dessel, P., & Brass, M. (2013). The influence of high-level beliefs on self-regulatory engagement: evidence from thermal pain stimulation. *Frontiers in Psychology*, *4*(September), 614.

Marchetti, C., & Della Sala, S. (1998). Disentangling the Alien and Anarchic Hand. *Cognitive Neuropsychiatry*, *3*(3), 191 – 207.

Mecklinger, A. D., von Cramon, Y. D., Springer, A., & Matthes – von Cramon, G. (1999). Executive control functions in task switching: evidence from brain injured patients. *Journal of Clinical and Experimental Neuropsychology, 21,* 606 – 619.

Miller, W. R., & Brown, J. M. (1991). Self-regulation as a conceptual basis for the prevention and treatment of addictive behaviors. In N. Heather, W. R. Miller, & J. Greely (Eds.), *Self-control and the addictive behaviors* (pp. 3 – 79). Sydney: Maxwell Macmillan.

O'Connor, D. A., Upton, D. J., Moore, J., & Hester, R. (2014). Motivationally significant self-control: Enhanced action withholding involves the right inferior frontal junction. *Journal of Cognitive Neuroscience*, 1 – 12.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Schel, M. A., Kühn, S., Brass, M., Haggard, P., Ridderinkhof, K. R., & Crone, E. A. (2014). Neural correlates of intentional and stimulus-driven inhibition: a comparison. *Frontiers in Human Neuroscience*, *8*(February), 27.

Schmeichel, B. J. & Zell, A. (2007). Trait self-control predicts performance on behavioral tests of self-control. *Journal of Personality, 75*(4), 743 – 756.

Smith, J. L., Mattick, R. P., Jamadar, S. D., & Iredale, J. M. (2014). Deficits in behavioural inhibition in substance abuse and addiction: A meta-analysis. *Drug and Alcohol Dependence*.

Swick, D., Ashley, V., & Turken, U. (2008). Left inferior frontal gyrus is critical for response inhibition. *BMC Neuroscience,* 9:102.

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers*.

Tyler, J. M., & Burns, K. C. (2008). After Depletion: The Replenishment of the Self's Regulatory Resources. *Self and Identity*.

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychol. Rev.*, *120*, 329 – 355.

Wessel, J. R. & Aron, A. R. (2014). Inhibitory motor control based on complex stopping goals relies on the same brain network as simple stopping. *NeuroImage, 103,* 225 – 234.

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., Middleton, K. R., Wing, R. R. (2014). Failure to replicate depletion of self-control. *PLoS ONE, 9*(10), e109950.

CHAPTER 4

The influence of high-level beliefs on self-regulatory engagement: Evidence from thermal pain stimulation[1]

Determinist beliefs have been shown to impact basic motor preparation, prosocial behavior, performance monitoring, and voluntary inhibition, presumably by diminishing the recruitment of cognitive resources for self-regulation. We sought to support and extend previous findings by applying a belief manipulation to a novel inhibition paradigm that requires participants to occasionally suppress a prepotent withdrawal reaction from a strong aversive stimulus. Our results suggest that reduction of free will beliefs lead to a form of intentional disengagement that influences action selection and inhibition. It is likely that disbelief in free will encourages participants to be more passive, to exhibit a reduction in intentional engagement, and to be disinclined to adapt their behavior to contextual needs.

[1] Lynn, M. T., Van Dessel, P., and Brass, M. (2013). The influence of high-level beliefs on self-regulatory engagement: Evidence from thermal pain stimulation. *Frontiers in Psychology, 4:* 614. doi: 10.3389/fpsyg.2013.00614

INTRODUCTION

The question of whether free will truly exists is an age-old philosophical question, tackled by thinkers ranging from Democritus to Russell. Yet most contemporary scientists have avoided the metaphysical and existential hurdles of free will, and instead investigate its impact on human action: how this phenomenon arises in the mind, and to what extent deterministic beliefs have an effect on our behavior (e.g., Wegner, 2003; Vohs & Schooler, 2008; Baumeister, Masicampo, & DeWall, 2009; Rigoni et al., 2011, 2012, 2013). The sensation of control over one's actions is an undeniably ubiquitous feature of human experience. People tend to believe they are responsible for a given action if the causal principles of *consistency, priority*, and *exclusivity* are satisfied; that is, if their intentions are consistent with and experienced at a suitable interval prior to the relevant action, and there is no other reasonable explanation for the action arising (Wegner, 2003). Perception of personal control is further considered to be intrinsic, biologically necessary, and protective against environmental stressors (Leotti, Iyengar, & Ochsner, 2010).

Social psychological research has recently investigated the degradation of behavioral and social effects thought to follow from a belief in determinism. For instance, Vohs and Schooler (2008) found that inducing disbelief in free will, via reading of a determinist essay or series of statements, elicited an increase in cheating on the part of participants. In comparison with control subjects, anti-free will participants in this case paid themselves a statistically improbable amount of money for performance on a problem-solving task, and more frequently permitted themselves to view answers when given the opportunity to cheat. Under similar conditions, Baumeister et al. (2009) found that participants with weakened free-will beliefs showed increased aggression and decreased helping behavior. Likewise, an increase in mindless conformity and a decrease in counterfactual thinking, assumed to be adaptive for learning and social adaptation, have been reported to accompany deterministic beliefs (Baumeister, Crescioni, & Alquist, 2011; Alquist, Ainsworth, & Baumeister, 2013). Interestingly, when these studies included a condition promoting free will, results were consistent with the control group, suggesting that a belief in free will is a common default state.

More recent research in the domain of Cognitive Psychology has revealed an impact of deterministic beliefs even on basic levels of motor control. Rigoni et al. (2011)

used a manipulation identical to that of Vohs and Schooler (2008, Experiment 1) to alter participants' belief in free will. They observed that participants who were induced to disbelieve in free will showed reduced amplitudes of the readiness potential, an electrophysiological marker of unconscious motor preparation (Rigoni et al., 2011). In a subsequent study (Rigoni et al., 2013) it was found that performance monitoring, as indicated by post-error slowing, was also diminished in participants induced to disbelieve in free will. This may indicate a reduction in the recruitment of self-regulatory processes, and less inclination to adjust one's behavior according to circumstantial needs, on the part of anti-free will participants.

Finally, this belief manipulation has been applied to an important facet of self-control, namely *intentional inhibition*, or the ability to voluntarily suppress a prepotent action plan (Brass & Haggard, 2007). The study in question (Rigoni et al., 2012) employed a task developed by Kühn, Haggard, and Brass (2009) that overcame a limitation of the well-supported literature on externally-generated stopping (see Aron, 2007, for a review) by enabling voluntary choice behavior to be experimentally investigated within an inhibition paradigm. In this task, participants were occasionally asked to freely decide whether to stop a prepared action (button pressing to halt the progress of a marble rolling down a ramp). Both intentional inhibition and perceived self-control were shown to be adversely affected by an anti-free will manipulation (Rigoni et al., 2012). These findings were interpreted such that weakened free will beliefs lead to a reduction in intentional effort, which then causes participants to select the less demanding response option (in this case to execute the pre-planned response).

The goal of the present study was to support and extend prior research on the influence of free will beliefs upon intentional inhibition, by investigating whether inducing determinist beliefs might in turn influence one's intentional engagement in self-regulatory behavior. However, while previous studies have investigated intentional inhibition in rather artificial experimental situations in which participants have hardly any prior motivation to act or inhibit, we sought to address voluntary inhibition in a more ecologically valid setting in which behavioral urges are present. To this end, our secondary goal was to develop and pilot a novel experimental paradigm for disentangling intentional from instructed inhibition.

Pain was selected as the behaviorally relevant stimulus for our purposes. Management of the pain avoidance response can be seen as a compelling component of the affective response system; the organism is strongly motivated to avoid the pain sensation (Campbell & Misanin, 1969; Elliot, 2006). We can therefore consider management of this urge as a window into how we suppress our most basic drives, and a classical instance of self-control. The pain avoidance response can of course be highly automatized, for instance when one reflexively jerks their hand away from a hot stove. However, at times other goals call for self-control to be exerted for the suppression of this avoidant urge, such as when the heat comes not from the stove, but from a plate of food. In this case, one might choose to suppress the highly prepotent reaction momentarily in favor of satisfying the opposing basic urge of hunger (cf. Morsella, 2005).

Our paradigm required participants to occasionally inhibit a prepotent withdrawal reaction from a heat source applied to their inner wrists. In half the trials, participants were able to choose whether to inhibit the withdrawal response or to immediately terminate the trial. The advantage of this manipulation is that it requires strong (and consistent; the urge to withdraw does not fade) self-control to withstand the thermal pain. In that sense, it is in stark contrast to standard laboratory tasks involving self-regulation and agency. The design also ensures that acting and inhibiting were equally distributed in the non-choice, or directed, trials, thereby discouraging any response bias and ensuring a comparable number of trial in each design cell. To manipulate free will beliefs, we used a Velten procedure (Velten, 1968) similar to that used in previous experiments (Vohs & Schooler, 2008, Experiment 2; Baumeister et al., 2009), in which participants are required to read and reflect upon a series of statements (see Appendix 1 for a complete list). Immediately prior to each trial, participants were presented with a statement and asked to retain the statement in memory until the end of the block. Statements were either neutral or meant to induce anti-free will beliefs (between-subjects). These statements were shown during the inter-trial interval in order to reduce potential pain preparation and decision-making strategies. We hypothesized that inducing disbelief in free will would lead participants to exhibit a reduction in intentional engagement, to lack adaptive strategies, and to be disinclined to adapt their behavior to contextual needs.

METHODS AND MATERIALS

Participants

Fifty-four Dutch-speaking undergraduate students enrolled in the study; all gave written consent prior to participation. They received either course credit or a payment of ten euros for their participation. All participants had normal or corrected-to-normal vision and reported no neurological deficits. The study was conducted in accordance with the Declaration of Helsinki, and the approval of Ghent University's Ethical Committee was obtained in advance. After determining participants' individual pain thresholds, those who did not report sufficient pain (i.e., their threshold surpassed fifty degrees – beyond the safety limitations of the stimulating equipment) were removed from the study. A total of 48 participants (12 male, tested individually) completed the entire experiment.

Procedure

Threshold determination. Pain was induced via a thermode connected to a Medoc PATHWAY device (MEDOC, Haifa, Israel), an apparatus designed to induce thermal pain using cold or hot stimulation. The threshold at which participants felt sufficient pain was determined by exposing each participant to 26 trials in which the thermal sensation gradually increased over five seconds from 32°C to a randomized destination temperature between 45 and 50°C (in increments of .25 degrees), a slope comparable to the experimental trials. After each trial, the thermode returned instantly to baseline temperature, and participants were asked to rate their perceived pain on a scale from zero to eight, with zero being no pain and eight being the worst possible pain. The destination temperature employed in the main experiment was computed for each participant as the highest temperature at which they rated their pain as a six. This method was revealed during piloting to yield more accurate tolerance threshold measurements than merely requiring participants to indicate the maximum heat they could withstand when exposed to a steadily increasing temperature. Importantly, participants were free to press a button at any point during the threshold determination in order to terminate the trial.

***Task and Stimuli*.** Participants received painful heat stimulation during each trial, applied via a thermode to alternating inner wrists. The images of three geometric

shapes (triangle, square, circle) were used as cues to indicate the trial type. Depending on the cue, participants were requested to either press the button as quickly as possible ('directed action,' 25% of trials), inhibit this response and endure the pain ('directed inhibition,' 25% of trials), or make a voluntary decision to either button press immediately or persist until the end of the trial ('choice,' 50% of trials). In the latter case, participants were requested to make their choices approximately equal over the course of the experiment, but not to use any particular strategies or to decide in advance of the presentation of the cue. In a practice block, absent pain stimulation, participants were trained on the cues. A pilot study had revealed that participants are typically around 200 ms slower to respond on choice action trials than on directed action trials, reflecting the additional time needed for the choice decision. Accordingly, to make stimulation as identical as possible across action conditions, 200 ms of thermal stimulation was added to directed action trials, following the button press.

Each trial was preceded by a statement ('neutral' or 'anti-free will,' see below) with a duration of twelve seconds. After a delay of one second, a fixation cross was presented and the temperature of the thermode began to gradually increase from a baseline of 32°C to the participant's individually determined threshold. After five seconds, one of the three task cues appeared in place of the fixation cross. The temperature remained at threshold for the next 2 seconds, or until the participant pressed the button to terminate both the pain stimulation and the trial. Afterwards, prompts for ratings of the perceived pain and 'urge to terminate the trial by pressing the button' (both on a scale of 0 – 8) remained on screen until participants responded. Participants were then cued to alternate the arm placed atop the thermode. The arm not being stimulated was used to button press (thereby providing a response time for action trials) and was placed atop the opposing wrist, in order to lend weight and make it more difficult for participants to inadvertently withdraw from pain rather than button pressing. A schematic overview of a possible trial in the anti-free will condition is presented in Figure 1.

The assignment of geometric shapes to trial types, and the order of the first-stimulated wrist were counterbalanced across subjects. Each participant had to perform 120 trials in total, being divided into six blocks of twenty trials presented in randomized sequence. In each block, participants were given ten trials in which they were cued to

make a decision, five trials in which they were cued to push and five trials in which they were cued to inhibit their withdrawal response. Importantly, participants were free to press a button to immediately terminate the thermal sensation at any point during the experiment.

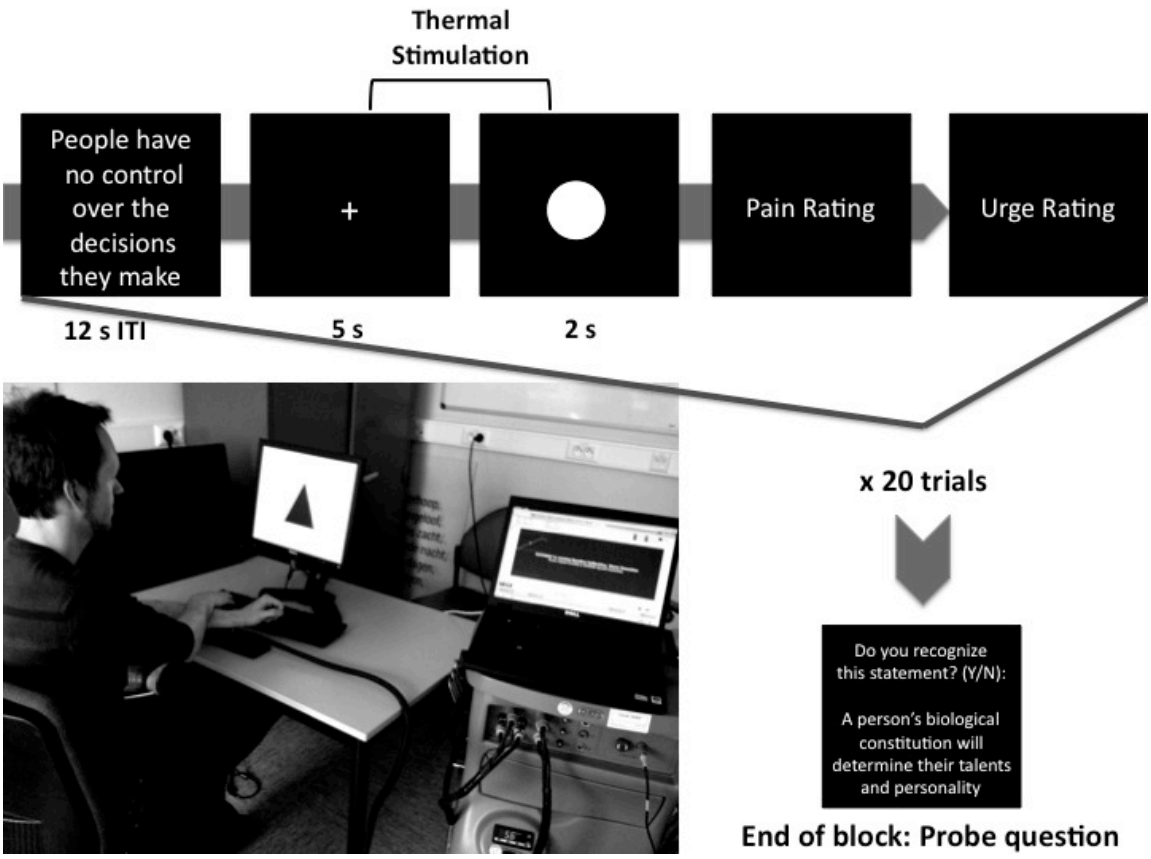

Figure 1. ***Schematic overview of a sample block (Anti-free will condition). Note that there was no time limit for pain and urge rating responses.***

Manipulation of free will beliefs. Participants were randomly assigned to either the control condition or the anti-free will condition (24 in each condition). All participants were required to read discrete statements presented on-screen during the inter-trial interval. They were instructed to retain this information until the end of the block, at which point a probe question concerning statement recognition was presented on the screen (see Appendix 2). The probe questions were inserted to verify that participants had attended to the statements as directed, and to support a cover story that the study's goal was to test the influence of pain on memory. After feedback on

the accuracy of their answer was given, a novel set of statements was presented, and subjects were instructed to remember these subsequent statements instead. The statements were either neutral or designed to tap into free will beliefs, with 60 unique statements in each group. Over the course of the experiment, control participants were exposed to each of the 60 neutral statements twice, while participants in the anti-free will condition were shown each of the 60 statements related to free will beliefs twice. Furthermore, in the anti-free will condition, the three trial types (directed action, directed inhibition, choice) were divided equally over each of the three statement categories.

A total of 90 statements were collected from a variety of questionnaires and articles involving free will beliefs (e.g., Vohs & Schooler, 2008; Paulhus & Carey, 2011; Carey, 2005), or were produced based on these inventories. These 90 statements were selected with the aim of being related to certain aspects of free will beliefs; thirty statements were related to the idea that people do not have a free will (e.g., "scientists tell us that people have no free will"), thirty statements concerned beliefs in scientific determinism (e.g., "the environment someone is raised in determines their success as an adult") and thirty statements were related to beliefs in fatalistic determinism (e.g., "you can't change your destiny, no matter how hard you try"). Another 90 neutral statements were selected, stating facts and ideas that were unrelated to beliefs in free will (e.g., "an ostrich's eye is bigger than its brain").

The combined 180 statements were then rated online (http://www.thesistools.com) by 38 participants, none of whom participated in the main experiment. Participants rated how difficult they would find the statement to recall, and the degree to which the statement was in line with either a disbelief in free will, a belief in scientific determinism, or a belief in fatalistic determinism. These questions were based on the factors laid out by Paulhus and Carey (2011) and were expressed in layman's terms for ease of understanding.

A total of 120 statements were selected based on the ratings drawn from this pre-test. The twenty statements that had received the highest ratings in each belief category were chosen, for a total of sixty experimental statements. Sixty neutral statements were matched for difficulty with these statements. Crucially, the

experimental statements and the control statements did not differ with regard to their difficulty to recall (experimental: M = 1.59; neutral: M = 1.60), $t(7) = 0.86$, $p = .82$.

Questionnaires. Two days prior to their participation in this study, participants completed an array of questionnaires concerning memory, anxiety and free will beliefs. Questions about memory and anxiety were inserted to support the aforementioned cover story. Questions regarding free will beliefs consisted of the entire battery of the Free Will and Determinism questionnaire (FAD-Plus, Paulhus & Carey, 2011). Following the experimental session, participants were requested to complete the FAD-Plus questionnaire a second time to determine whether or not the experimental statements had an effect on the relevant belief system.

RESULTS

Manipulation check

To test the effectiveness of the belief manipulation, a mixed design ANOVA was conducted on participants' total FAD-scores before and after the experiment using Time (Pre-test vs. Post-test) as a within-subject factor and Belief condition (Anti-free will vs. Control) as a between-subjects factor. Total FAD-scores were calculated for each participant such that higher values indicate less belief in free will, by reverse scoring the Free Will subscale and combining it with the other three subscales (Scientific Determinism, Fatalistic Determinism, and Unpredictability). The analysis revealed a significant interaction between Time and Belief Condition, $F(1,46) = 4.19$; $p < .05$ (Figure 2), such that participants in the experimental condition scored significantly higher after the experiment than before (Post-test: M = 80.0, SD = 8.9; Pre-test: M = 76.3, SD = 8.5), $t(23) = 3.23$, $p < .01$, indicating a weakening of beliefs in free will. No such effect was observed for participants in the control condition (Post-test: M = 76.9, SD = 8.9; Pre-test: M = 76.6, SD = 9.4), $t(23) = 0.29$, $p = .78$.

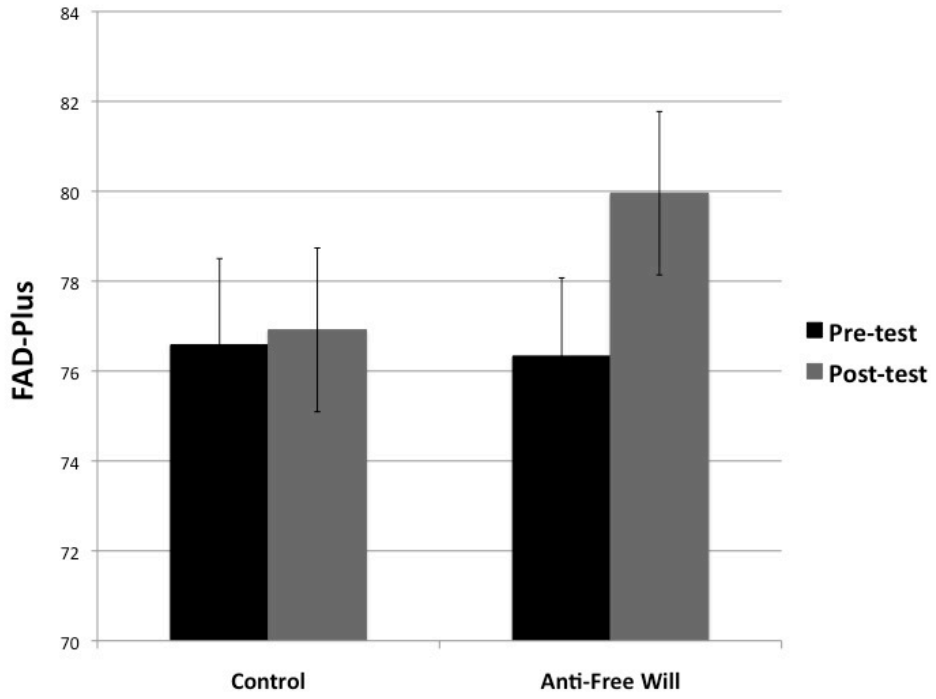

Figure 2. ***Mean total scores on the FAD-Plus questionnaire as a function of Belief condition (Control vs. Anti-free will) and Time (Pre-test vs. Post-test). Higher scores indicate increased disbelief in free will.***

Data preparation

Despite efforts towards optimizing the pain threshold procedure, the grand mean pain rating across participants was rather low (*M* = 4.6; *SD* = 1.11). Crucially, in the debriefing questionnaire, more than half (N = 26) of all participants stated that they had not needed to exert any effort to withhold the pain-withdrawal response during the experiment. As pain is a key factor in this experiment, we decided to restrict our analyses to participants that reported a sufficient level of pain throughout the whole of the experiment. We therefore excluded all participants with mean pain ratings lower than the median of the subjective pain scale, namely 4.5. All further analyses were performed on this subset of 25 'high pain' participants (8 male): 12 participants in the anti-free will condition and 13 participants in the control condition. Results for the excluded 'low pain' participants may be found in Appendix 3.

Behavioral analyses

Between-group means and standard deviations are reported in Table 1.

Table 1. Between-group means and standard deviations

	Control	Anti-Free Will
	Mean (SD)	Mean (SD)
Reaction Times (ms)		
All Action Trials	658 (124)	736 (101)
Choice Action Trials	748 (162)	871 (133)
Directed Action Trials	552 (121)	582 (94)
Proportion Inhibition (%)	40.59 (9.64)	42.43 (10.22)
Pain Ratings (across trials)	5.5 (0.9)	5.4 (0.6)
Urge Ratings		
Across Trials	4.4 (1.3)	4.7 (1.5)
Choice Trials	4.5 (0.4)	4.5 (0.4)
Directed Trials	4.3 (0.4)	4.8 (0.3)

Reaction times. On trials in which participants were cued to button press, participants performed the correct response in nearly all trials (M = 99%, SD = 2%). We expected anti-free will participants to be significantly slower than controls, particularly on choice trials. A mixed design ANOVA on RTs, with Instruction (Choice vs. Directed) as a within-subjects factor and Belief condition (Anti-free will vs. Control) as a between-subjects factor, revealed a main effect of Instruction, $F(1,23) = 79.310$, $p < .01$, such that participants were slower to respond on choice trials (Choice: M = 807 ms, SD = 158 ms; Directed: M = 567 ms, SD = 108 ms), consistent with piloting and reflecting the time needed for a response decision. A main effect of Belief condition revealed a non-significant trend, $F(1,23) = 2.958$, $p = .099$, indicating that anti-free will participants tended to be slower to respond than controls (though this interpretation should be approached with caution due to the marginal significance level). Further, the interaction between Instruction and Belief condition trended towards significance, $F(1, 23) = 2.928$, $p = .10$. Planned comparisons revealed an RT difference between anti-free will participants and controls on choice action trials, $t(23) = -2.07$, $p < .05$, Cohen's d =

.84 (Figure 3), such that anti-free will participants were significantly slower to respond when given a choice than were controls. No such effect was found on directed action trials, $t(23) = -.69$, $p = .497$, $d = .27$.

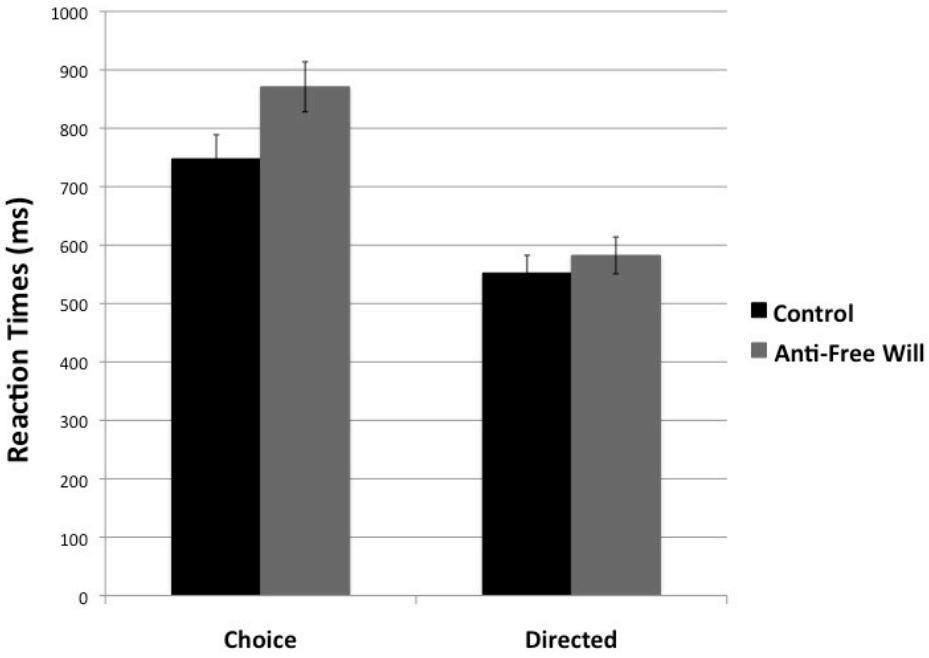

Figure 3. ***Reaction times on press trials, between-subjects. Values depicted are means and standard errors.***

Correlation of FAD difference scores with choice reaction times. To examine the relationship between participants' RTs and free will beliefs more thoroughly, we performed an additional correlation analysis. The aim of this analysis was to test to what extent the slowed responding on choice action trials was related to the effectiveness of the belief manipulation. To this end, we first computed each participant's change in anti-free will beliefs, across experimental condition (control participants were included to ensure sufficient variability), by subtracting participants' post-experimental scores on the anti-free will subscale of the FAD from their pre-experimental scores. Second, we computed a difference score of participants' mean RTs on choice and directed action trials to create an index of each participant's decision time at pushing the button. There was a significant positive correlation between the two difference scores, $r(23)=0.40$, $p < .05$ (Figure 4), reflecting that those subjects who

showed a stronger reduction in free will beliefs were also slower to make the decision to press the button.

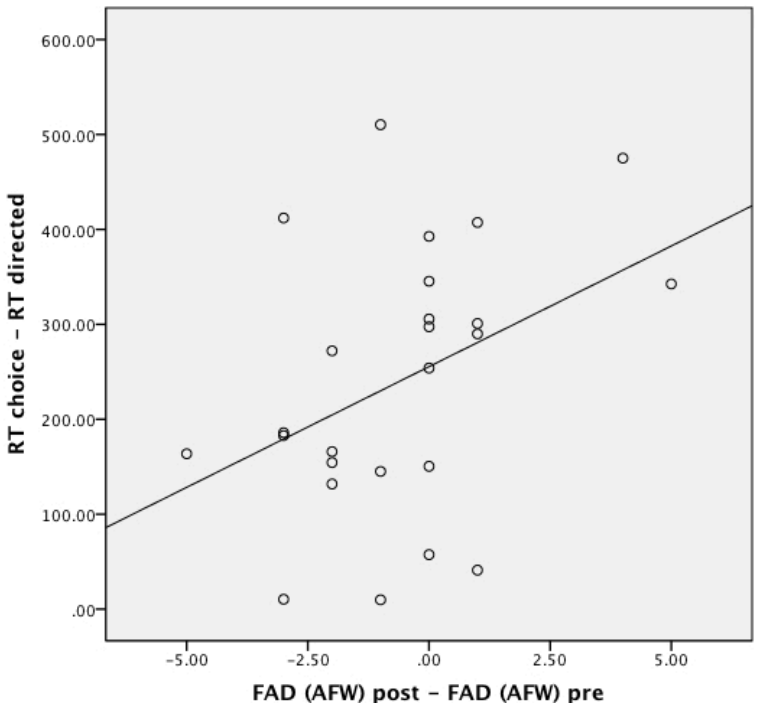

Figure 4. ***Correlation of difference scores (post-test minus pre-test) on the anti-free will subscale of the FAD-Plus with the decision response time index (mean response times on choice minus directed trials).***

Proportion of inhibition on choice trials. On trials in which participants were cued to choose between acting and inhibiting, participants opted to inhibit in 41.47% of all trials (*SD* = 9.76%). The proportion of inhibition on choice trials was analyzed in an independent-samples t-test, revealing no significant difference between anti-free will participants and controls, $t(23) = -.462$, $p = .648$. This lack of a difference between experimental groups, which is in contrast to the findings of Rigoni et al. (2012), may be due to the experimental design, which, unlike previous studies, discourages response biases by using an equal proportion of directed action and inhibition trials.

Ratings

Pain ratings. We began by computing pain ratings across all participants for the first and second halves of the experiment to ensure that participants did not adapt to the pain stimulation over the course of the experiment. No differences in pain ratings

were observed between the trials of the first and the second half of the experiment (First half: *M* = 5.4, *SD* = 0.8; Second half: *M* = 5.5, *SD* = 0.8), *t*(24) = -0.58, *p* = .57.

Participants reported a grand mean pain rating of 5.5 (*SD* = 0.74). Pain ratings were analyzed in a mixed design ANOVA using Belief condition as a between-subjects factor, and Response (Action vs. Inhibition) and Instruction (Directed vs. Choice) as within-subject factors. The main effect of Belief condition was not significant, *F*(1,23) = 0.13, *p* = .73, reflecting that subjective pain across trials was equivalent for the two groups. However, there was a significant main effect of Response (Action: *M* = 5.3, *SD* = 0.2; Inhibition: *M* = 5.7, *SD* = 0.1), *F*(1,23) = 12.60, *p* < .01, indicating higher perceived pain on inhibition compared with action trials, presumably due to the lengthier pain stimulation. Moreover, there was an interaction effect of Response x Instruction, *F*(1,23) = 7.94, *p* = .01, reflecting that inhibition trials were rated as less painful when they were voluntarily chosen rather than instructed (Choice: *M* = 5.5, *SD* = 0.8; Directed: *M* = 5.8, *SD* = 0.6), *t*(24) = 3.38, *p* < .01, while there was no such difference between chosen and directed action trials (Choice: *M* = 5.4, *SD* = 1.0; Directed: *M* = 5.2, *SD* = 0.9), *t*(24) = -1.54, *p*= .14. Importantly, the lack of a difference between the mean pain ratings of anti-free will and control participants suggests that our findings are not solely due to differences in the overall subjective experience of pain.

Urge ratings. Participants reported a grand mean urge rating of 4.5 (*SD* = 1.4). Urge ratings were analyzed with a mixed design ANOVA akin to that of the pain ratings. The analysis revealed a significant main effect of response, reflecting greater urges on action trials (Action: *M* = 4.8, *SD* = 0.3; Inhibition: *M* = 4.2, *SD* = 0.3), *F*(1,23) = 4.98, *p* < .05. There was also a significant interaction effect of Response x Instruction, *F*(1,23) = 6.49, *p* < .05. Consistent with the pain ratings, participants reported a reduced urge on choice compared with directed inhibition trials (Choice: *M* = 4.0, *SD* = 1.6; Directed: *M* = 4.5, *SD* = 1.7), *t*(24) = 2.67, *p* < .05, while there was no such difference between choice and directed action trials (Choice: *M* = 5.0, *SD* = 1.4; Directed: *M* = 4.6, *SD* = 1.6), *t*(24) = -1.70, *p* = .10. The main effect of Belief condition was not significant, *F*(1,23) = 0.10 , *p* = .76. Crucially however, there was a significant interaction effect of Belief condition x Instruction, *F*(1,23) = 6.22, *p* < .05. Post-hoc t-tests revealed that participants in the anti-free will condition tended to report a stronger urge to press on directed trials than on choice trials, *t*(11) = 2.044, *p* = .066, whereas this was not the case for control

subjects, $t(12)= -1.465$, $p=.17$ (Figure 5). This may be indicative of a greater urge to act when externally instructed on the part of anti-free will participants. Similar results were obtained by Alquist et al. (2013), who found that anti-free will participants conformed more to external pressure.

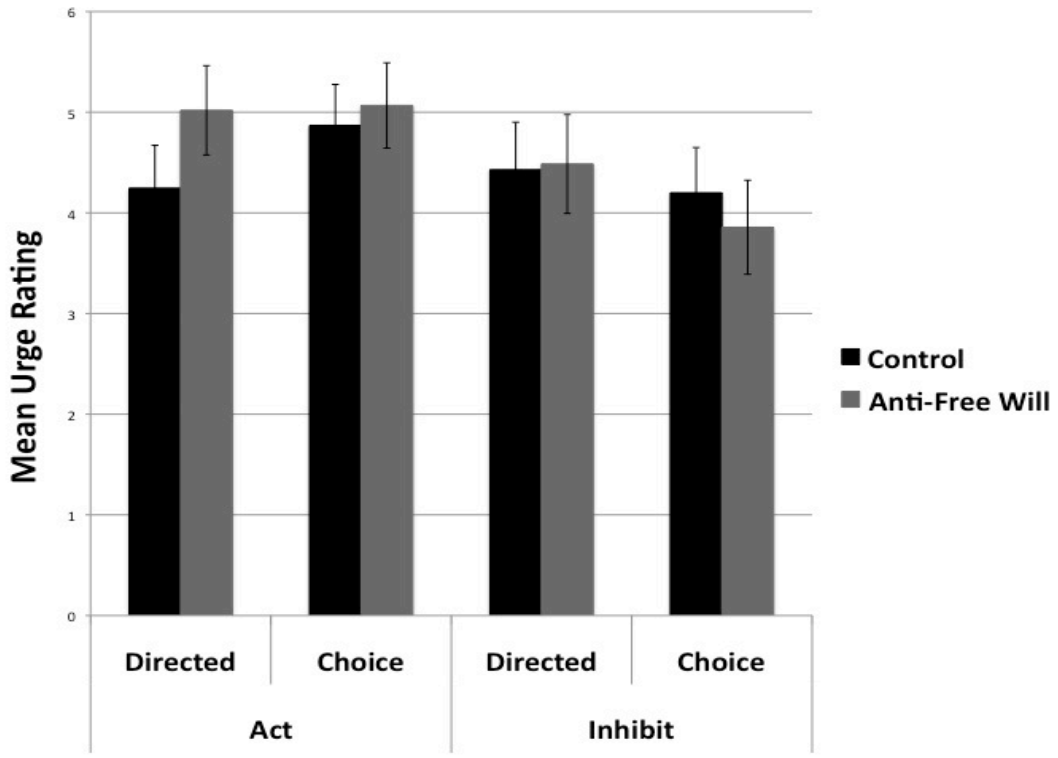

Figure 5. ***Urge ratings as a function of Instruction (Choice vs. Directed) and Belief condition (Control vs. Anti-free will). Values depicted are means and standard errors.***

Adaptive strategies on choice trials

Based on the hypothesis that anti-free will participants might lack adaptive strategies, we conducted an exploratory analysis in which we investigated whether preceding trial pain or trial type had an influence on response selection during choice trials. We assumed that high pain trials might create a strong incentive to 'quit' when subsequently given a choice, thereby activating a strategy that is protective of the organism. Similarly, participants might attempt to create subjectively easier response sequences when granted the opportunity. These strategies would presumably only be

present for control participants, as anti-free will participants tend to be less inclined to adjust their behavior to the present situation (Rigoni et al., 2013).

Pain on preceding trial. To investigate the influence of pain on subsequent choice behavior, we computed each participant's mean pain rating for the trials preceding choice inhibition and choice action trials. A mixed design ANOVA with factors of Belief condition (Anti-free will vs. Control) and Response (Choice Action vs. Choice Inhibition) was then conducted on mean pain rating for n-1 trials. The analysis revealed no main effects or interactions, *F*s < .838, *p*s > .36, indicating that pain ratings on the preceding trial did not differ between choice inhibition and choice action trials, for either experimental group. This would suggest that participants do not use recent pain as a factor in deciding whether to act or inhibit when given the choice.

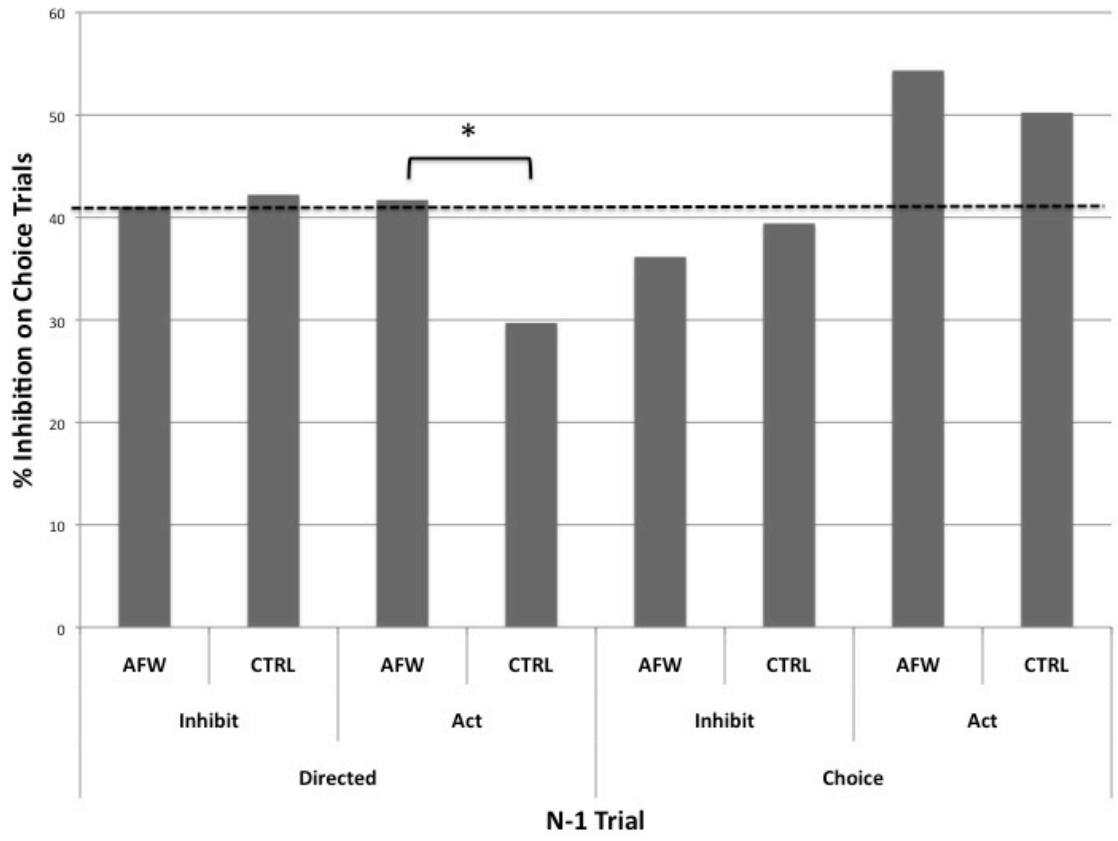

Figure 6. ***N-1 trial contribution to response tendencies in each experimental group. Compared with anti-free will participants (AFW), control participants (CTRL) tend to inhibit less often following a directed press trial. The dashed line indicates the grand mean proportion of inhibition. * p < 0.05.***

Response styles. To investigate response styles, we computed mean proportions of inhibition during choice trials following each of the four trial types. A mixed design ANOVA with factors of Belief condition (Anti-free will vs. Control), n-1 Instruction (Choice vs. Directed), and n-1 Response (Action vs. Inhibition) was then conducted on mean proportion of inhibition in choice trials. This gave an index of how often participants chose to inhibit rather than act following a particular trial type (Figure 6). The analysis revealed a main effect of n-1 Instruction (Choice: M = 45.0% inhibition on subsequent choice trial; Directed: *M* = 38.7% inhibition on subsequent choice trial), $F(1,23) = 6.366$, $p < .05$, such that participants tended to choose to inhibit more often following a choice trial. There was also a significant interaction between n-1 Instruction and n-1 Response, $F(1,23) = 11.460$, $p < .01$, such that participants chose to inhibit more often following a choice action trial (M = 52.2%) than any other trial type (Choice Inhibit n-1 = 37.9%; Directed Action n-1 = 35.5%; Directed Inhibition n-1 = 41.7%), *t*s > 2.64, *p*s < .05. Furthermore, there was a non-significant trend towards an interaction between n-1 Response and Belief condition, $F(1,23) = 3.523$, $p = .07$. Anti-free will participants tended to inhibit more often following an action trial (*M* = 48.0%) than an inhibition trial (*M* = 38.6%), $t(11) = -2.164$, $p = .05$, $d = .63$, whereas this was not the case for controls (Action n-1: *M* = 40.0%; Inhibition n-1: *M* = 40.8%), $t(12) = .251$, $p = .806$, $d = .03$. This may indicate a more explicit tendency to alternate in an attempt to satisfy the 50% choice instruction. Finally, post-hoc t-tests confirmed that the primary difference in proportion of inhibition between experimental groups lay in directed action n-1 trials. Control subjects chose to inhibit significantly less often than anti-free will participants following a directed action trial (Control: *M* = 29.7%; Anti-free will: *M = 41.7%*), $t(23) = -2.490$, $p < .05$, $d = .99$. This may be indicative of an additional adaptive strategy on the part of control participants, as response repetitions are subjectively less effortful than response switches.

DISCUSSION

In the present study, we employed a novel experimental approach using thermal pain stimulation in order to demonstrate the moderating nature of high-level beliefs on self-regulation. In particular, we sought to probe whether reducing participants' belief in free will could lead to a form of intentional disengagement that influences selection and inhibition of action within a 'hot' motivational system (Metcalfe & Mischel, 1999).

In line with our predictions, participants who were induced to disbelieve in free will were significantly slower to initiate a response on trials in which they chose to act. This directly corresponds to the hypothesis that anti-free will participants would exhibit less intentional engagement. Interestingly, this effect is only evident when a response has to be executed internally rather than externally, suggesting not a global passivity, but rather a specific impairment in voluntary self-regulation. The amount of slowing on choice action trials was furthermore correlated with the degree of the effectiveness of the belief manipulation, suggesting a direct link between the weakening of free will beliefs and intentional behavioral engagement. This mirrors the finding by Rigoni et al. (2011) in which decreases in the readiness potential were correlated with a change in anti-free will scores. Moreover, anti-free will participants reported greater urges to terminate the trial when their behavior was guided by the cue compared to when they were able to freely choose, suggesting a disengagement from the task when externally instructed. Importantly, and in contrast with previous studies, the aforementioned differences are not confounded by differential response biases, as the proportion of inhibition in choice trials was equivalent between control and anti-free will participants.

Our analysis of potentially adaptive strategies revealed surprising results. Participants do not appear to use recent pain as a criterion in deciding whether to act or inhibit when given the choice. However, we do find differences between the experimental groups in terms of their response styles. Interpretations are merely speculative at this point, but one could suppose that control participants select a subjectively easier strategy when exhibiting a bias to repeat an action response. On the other hand, one could interpret the anti-free will participants as selecting the less effortful strategy, by avoiding two (subjectively more painful) inhibition trials in a row. In the future, this could be disentangled by presenting blocks composed solely of choice trials in order to determine, via longer choice trial sequences, which is the favored strategy: response repetitions or avoidance of effortful combinations.

Taken together, the present study supports and extends previous research on intentional inhibition. In particular, it is the first to investigate voluntary inhibition of behavior in an ecologically valid experimental setting that involves hot motivational systems. Participants reported less pain and a reduced urge to terminate the trial on choice inhibition trials compared with directed inhibition trials, while choice and

directed press trials were more comparable. Thus the pain paradigm we introduce offers an effective way to dissociate between voluntary and instructed inhibition on a behavioral level, which opens the door to new ways of investigating inhibition in which behaviorally-relevant options are available to the participant.

That being said, as this study served as a first pilot of a novel paradigm, our investigation must be seen as exploratory in nature, and our conclusions considered accordingly. The exclusion of participants who did not experience sufficient pain levels is an unfortunate limitation of the present line of research. Future studies should endeavor to ensure that a sufficient pain tolerance threshold is obtained for each participant, or that unsuitable participants are excluded in advance of testing. This may require rigorous pre-testing of criteria such as whether participants are able to reliably report their tolerance thresholds, and whether or not they adapt too quickly to pain over the course of the experiment.

On a larger scale, the observed effects also exemplify a growing body of research that reveals the influence of higher-order beliefs and metacognitions on behavioral control. As discussed earlier, determinist beliefs have been shown to have an effect on prosocial behavior (Vohs & Schooler, 2008; Baumeister et al., 2009, 2011), basic motor and cognitive processes (Rigoni et al., 2011, 2013), intentional inhibition (Rigoni et al., 2012), and now on self-regulation of a 'hot' incentive response system (Morsella, 2005). Yet free will beliefs are not the only higher-order cognitions capable of influencing a variety of processes underlying behavioral control.

For instance, one factor that has been proposed to have a strong influence on self-control is 'ego depletion,' or the phenomenon in which exertion of self-control exhausts a common regulatory resource, leading to hindered performance on subsequent tasks (Muraven, Tice, & Baumeister, 1998; Vohs et al., 2008; Baumeister, 2009; Hagger et al., 2010). However, recent research has revealed that participants' relevant belief systems are likely to be more crucial than actual depletion when it comes to self-regulatory capacity. For instance, Job, Dweck, & Walton (2010) demonstrated that only participants who thought of willpower as a limited resource demonstrated the typical pattern of ego depletion, while the effect was completely absent in participants who lacked this conviction. Similarly, Clarkson and colleagues (2010) found that regardless of how depleted participants actually were, if they

perceived themselves as less depleted, they failed to demonstrate ego depletion effects during subsequent task performance (see also Vohs, Baumeister, & Schmeichel, 2012). These observations indicate that beliefs regarding regulatory resources are distinct from the resources themselves, and can impact task performance independently, thus mirroring our finding of reduced engagement in self-control on the part of subjects whose belief in free will has been diminished.

A similar influence of high-level beliefs has been observed for intentional binding, i.e., an implicit measure of self-agency in which the perception of voluntary movement is shifted forward in time towards a subsequent sensory effect, and vice versa. Desantis et al. (2011) demonstrated that intentional binding was stronger when participants were led to believe they were responsible for the action effect than when they believed that another participant was the causal agent. This finding indicates that high-level beliefs about authorship can influence low-level sensorimotor processes, and thus bias the experience of self-agency. A similar mechanism may also contribute to our findings, as belief in the concept of free will has been shown to be associated with the experience of self-agency (Aarts & van den Bos, 2011). Thus, the weakening of free will beliefs may have discouraged participants from engaging in self-regulatory behavior due to the subjective feeling of reduced control over the outcomes of their actions. Future research should directly address this question.

The present study demonstrates the manner in which a disbelief in free will may lead participants to be less inclined to put effort into self-regulatory behavior, and to rely on the external environment rather than internally-determined strategies, even in the face of a strong incentive stimulus. It seems, therefore, that metacognitions have a nontrivial impact on behavioral control. Furthermore, a fundamental belief in control over one's actions may prove to be an integral prerequisite for self-regulatory investments. Future studies should more directly investigate the mechanisms by which higher-order beliefs impact the recruitment of self-control.

Acknowledgements

This work was supported by the European Science Foundation's EUROVETO project (09-ECRP-020).

REFERENCES

Aarts, H. & van den Bos, K. (2011). On the foundation of beliefs in free will: Intentional binding and unconscious priming in self-agency. *Psychological Science, 22*(4), 532-537.

Alquist, J., Ainsworth, S., & Baumeister, R. (2013). Determined to conform: Disbelief in free will increases conformity. *Journal of Experimental Social Psychology, 49*, 80-86.

Aron, A. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist, 13*, 214-228.

Baumeister, R.F., Crescioni, A.W., & Alquist, J.L. (2011). Free Will as advanced action control for human social life and culture. *Neuroethics, 4*, 1-11.

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Pro-social benefits of feeling free: Manipulating disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin, 35*, 260-268.

Brass, M., & Haggard, P. (2007). To do or not to do: The neural signature of self-control. *Journal of Neuroscience, 27,* 9141–9145.

Carey, J.M. (2005). Development and validation of a measure of free will belief and its alternatives (Unpublished manuscript). The University of British Columbia, Vancouver.

Clarkson, J. J., Hirt, E. R., Jia, L. & Alexander, M. B. (2010). When perception is more than reality: The effects of perceived versus actual resource depletion on self-regulatory behavior. *Journal of Personality and Social Psychology, 98,* 29-46.

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition, 20*, 1211 – 1220.

Elliot, A. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion, 30,* 111-116.

Fujita, K., Trope, Y., Liberman, N., Levin-Sagi, M. (2006). Construal levels and self-control. *J Pers Soc Psychol, 90*(3), 351-367.

Hagger, M.S., Wood, C., Stiff, C., & Chatzisarantis, N.L.D. (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychological Bulletin, 136*(4), 495 – 525.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21,* 1686–1693.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: How the ‘‘veto-area’’ exerts control. *Human Brain Mapping, 30,* 2834–2843.

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

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106,* 3 – 19.

Morsella, E. (2005). The function of phenomenal states: Supramodular interaction theory. *Psychological Review, 112,* 1000 – 1021.

Muraven, M. Tice, D.M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*(3), 774 – 789.

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring lay beliefs regarding free will and related constructs*. Journal of Personality Assessment, 93*, 96-104.

Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition, 21*(3), 1482–1490.

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, 22,* 613-618.

Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition, 127,* 264-269.

Velten, E. (1968). A laboratory task for the induction of mood states. *Behavioral Research and Therapy, 6,* 607–617.

Vohs, K.D., Baumeister, R.F., Schmeichel, B.J., Twenge, J.M., Nelson, N.M., & Tice, D.M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94,* 883-893.

Vohs, K.D., Baumeister, R.F., & Schmeichel, B.J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology, 48,* 943-947.

Vohs, K.D., & Schooler, J. (2008). The value of believing in free will: Encouraging a belief in Determinism increases cheating. *Psychological Science, 19,* 49–54.

Appendices 1 and 2 can be found at:
http://www.frontiersin.org/Cognition/10.3389/fpsyg. 2013.00614/abstract

APPENDIX 3

Results Including Low Pain Participants

Overview

The 'low pain' group consisted of 23 participants (4 males), with 12 participants in the anti-free will condition and 11 in the control condition. Total FAD-scores did not differ significantly between the low and high pain group, t(46) = 0.77, p=.44. The grand mean pain rating for the low pain group was 3.7 (*SD* = 0.57). The below results include all participants, and treat Pain Group as a between-subjects factor.

Behavioral analyses

Reaction times. On trials in which participants were cued to button press, participants performed the correct response in nearly all trials (*M* = 98%, *SD* = 3%). We expected anti-free will participants to be significantly slower than controls, particularly on choice trials. A mixed design ANOVA on RTs with Instruction (Choice vs. Directed) as a within-subjects factor, and Belief condition (Anti-free will vs. Control) and Pain Group (High vs. Low) as between-subjects factors, revealed a main effect of Instruction, $F(1,44) = 139.31$, $p < .001$, such that participants were slower to respond on choice trials (Choice: *M* = 795 ms, *SD* = 21 ms; Directed: *M* = 563 ms, *SD* = 15 ms), consistent with piloting and reflecting the time needed for a response decision. Neither the main effect of Belief condition nor the main effect of Pain Group were significant, *p*s > .31**.** Crucially however, there was a significant three-way interaction effect of Instruction x Belief Condition x Pain Group, $F(1,44) = 4.43$, $p < .05$. Planned comparisons revealed a significant RT difference between anti-free will participants and controls on choice action trials in the high pain group, $t(23) = -2.07$, $p < .05$, such that anti-free will participants were significantly slower to respond when given a choice (*M* = 871 ms, *SD* = 133 ms) than controls (*M* = 748 ms, *SD* = 162 ms). No such effect was found on exogenous action trials, $t(23) = -.69$, $p = .497$ (Anti-free will: *M* = 582 ms, *SD* = 94 ms; Control: *M* = 552 ms, *SD* = 121 ms). In the low pain group there were no significant RT differences between the Belief Conditions, *p*s>.40.

Correlation of FAD difference scores with choice reaction times. To examine the relationship between participants' RTs and free will beliefs more thoroughly, we performed an additional correlation analysis. The aim of this analysis was to test to what extent the slowed responding on endogenous action trials was related to the effectiveness of the belief manipulation. To this end, we first computed each participant's change in anti-free will beliefs, across experimental condition, by subtracting participants' post-experimental scores on the anti-free will subscale of the FAD from their pre-experimental scores. Second, we computed a difference score of participants' mean RTs on endogenous and exogenous action trials to create an index of each participant's decision time at pushing the button. There was a significant positive correlation between the two difference scores for participants in the high pain group, $r(25)=0.40$, $p < .05$, reflecting that those subjects who showed a stronger reduction in free will beliefs were also slower to make the decision to press the button. In line with the previous analysis, no such effect was observed for participants in the low pain group, $r(23)=0.04$, $p = .87$.

Proportion of inhibition on choice trials. On trials in which participants were cued to choose between pressing and inhibiting, participants opted to inhibit in 44.92% of all trials (*SD* = 12.60%). The proportion of inhibition on choice trials was analyzed in an ANOVA with Belief condition and Pain Group as between-subjects factors. This analysis revealed a main effect of Pain Group, $F(1,44) = 3.87$, $p = .05$, such that participants in the high pain group chose to inhibit less often than participants in the low pain group (high pain: *M* = 41.6%, *SD* = 2.5%; low pain: *M* = 48.6%, *SD* = 2.6%). However, there was no interaction effect, nor a significant main effect of Belief Condition, ps>. 70. A planned comparison revealed no significant difference between anti-free will participants (*M* = 42.43%, *SD* = 10.22%) and controls (*M* = 40.59%, *SD* = 9.64%) in the high pain group, $t(23) = -.462$, $p = .65$. This lack of a difference between experimental groups, which is in contrast to the findings of Rigoni et al. (2012), may be due to the experimental design, which, unlike previous studies, discourages response biases by using an equal proportion of directed press and inhibit trials.

Ratings

Pain ratings. Pain ratings were computed for the first and second halves of the experiment to ensure that participants did not adapt to the pain stimulation over the

course of the experiment. No differences in pain ratings were observed between the trials of the first and the second half of the experiment (First half: *M* = 4.6, *SD* = 1.1; Second half: *M* = 4.6, *SD* = 1.2), *t*(47) = 0.21, *p* = .84. Pain ratings were analyzed in a mixed design ANOVA using Belief condition as a between-subjects factor, and Response (Action vs. Inhibition) and Instruction as within-subject factors. The main effect of Belief condition was not significant (Anti-free will: *M* = 4.5, *SD* = 1.1; Control: *M* = 4.7, *SD* = 1.2), *F*(1,46) = 0.23, *p* = .63, reflecting that subjective pain across trials was equivalent for the two groups. However, there was a significant main effect of Response (Action: *M* = 4.2, *SD* = 0.2; Inhibition: *M* = 5.0, *SD* = 0.1), *F*(1,46) = 51.81, *p* < .001, indicating higher perceived pain on inhibition compared with action trials, presumably due to the lengthier pain stimulation. Moreover, there was a main effect of instruction, *F*(1,46) = 4.77, *p* < .05. Participants had higher pain ratings on directed than on choice trials (Directed: *M* = 4.7, *SD* = 0.2; Choice: *M* = 4.6, *SD* = 0.2). We also observed an interaction effect of Response x Instruction, *F*(1,46) = 20.12, *p* < .001, reflecting that inhibition trials were rated as less painful when they were chosen rather than directed (Choice: *M* = 4.8, *SD* = 1.1; Directed: *M* = 5.2, *SD* = 0.9), *t*(47) = 5.78, *p* < .001, while action trials were rated as less painful when they were directed rather than chosen (Directed: *M* = 4.1, *SD* = 1.4; Choice: *M* = 4.3, *SD* = 1.4), *t*(47) = -2.30, *p* < .05. Importantly, the lack of a difference between the mean pain ratings of anti-free will and control participants suggests that our findings are not solely due to differences in the overall subjective experience of pain.

Urge ratings. Participants reported a grand mean urge rating of 3.6 (*SD* = 1.7). Participants' mean urge ratings were higher in the high pain group (*M* = 4.5, *SD* = 1.38) than in the low pain group (*M* = 2.6, *SD* = 1.37), *t*(46) = 4.94, p < .001. Urge ratings were analyzed with a mixed design ANOVA akin to that of pain ratings. The analysis revealed a significant main effect of response, reflecting greater urges on action trials than inhibition trials (Action: *M* = 4.0, *SD* = 0.3; Inhibition: *M* = 3.2, *SD* = 0.3), *F*(1,46) = 8.46, *p* < .05. There was also a significant interaction effect of Response x Instruction, *F*(1,46) = 9.74, *p* < .05. Consistent with the pain ratings, participants reported reduced urges on choice compared with directed inhibition trials (Choice: *M* = 3.0, *SD* = 1.7; Directed: *M* = 3.4, *SD* = 1.9), *t*(47) = 2.89, *p* < .05, while this effect was reversed for choice and directed action trials (Choice: *M* = 4.2, *SD* = 2.1; Directed: *M* = 3.7, *SD* = 2.2), *t*(47) = -

2.27, $p < .05$. The main effect of Belief condition was not significant (Anti-free will: M = 3.7, SD = 1.6; Control: M = 3.5, SD = 1.8), $F(1,46) = 0.14$, $p = .72$. Notably however, we observed a trend towards a three-way interaction effect of Belief condition x Response x Instruction, $F(1,23) = 3.36$, $p = .07$. Post-hoc t-tests revealed that participants in the control condition tended to report a stronger urge to press on choice action trials than on directed action trials (Choice Action: M = 4.3, SD = 2.1; Directed Action: M = 3.4, SD = 2.0), $t(23) = 4.32$, $p < .001$, whereas this was not the case for anti-free will subjects (Choice Action: M = 4.2, SD = 2.1; Directed Action: M = 4.1, SD = 2.4), $t(23)= 0.21$, $p=.84$. Though strong conclusions cannot be drawn from this non-significant result, it may be indicative of a greater urge to act when externally instructed on the part of anti-free will participants. Similar results were obtained by Alquist et al. (2013), who found that anti-free will participants conformed more to external pressure.

CHAPTER 5

Priming determinist beliefs diminishes implicit (but not explicit) components of self-agency[1]

Weakening belief in the concept of free will yields pronounced effects upon social behavior, typically promoting selfish and aggressive over pro-social and helping tendencies. Belief manipulations have furthermore been shown to modulate basic and unconscious processes involved in motor control and self-regulation. Yet, to date, it remains unclear how high-level beliefs can impact such a wide range of behaviors. Here, we tested the hypothesis that priming disbelief in free will diminishes the sense of agency, i.e., the intrinsic sensation of being in control of one's own actions. To this end, we measured participants' implicit and explicit self-agency under both anti-free will and control conditions. Priming disbelief in free will reduced implicit but not explicit components of agency. These findings suggest that free will beliefs have a causal impact on the pre-reflective feeling of being in control of one's actions, and solidify previous proposals that implicit and explicit agency components tap into distinct facets of action awareness.

[1] In collaboration with Paul S. Muhle-Karbe, Henk Aarts, and Marcel Brass (under revision, Frontiers in Psychology).

INTRODUCTION

The question of whether free will truly exists has fascinated philosophers, psychologists, and neuroscientists for centuries (e.g., Haggard, 2008; Libet et al., 1983; Wegner, 2004). Yet contemporary empirical research typically avoids the existential question itself, and instead focuses on more tangible research questions concerning the consequences of (dis)belief in free will, and its relation to agentic causation and volition.

Seminal studies in the domain of social psychology have shown that weakening belief in the concept of free will, via reading of essays or statements that promote a determinist perspective, can greatly impact participants' subsequent social behavior. For instance, Vohs and Schooler (2008) found that participants who were primed with disbelief in free will paid themselves a statistically improbable amount of money and took advantage of opportunities to cheat more often than a group of control participants who read texts unrelated to free will. Likewise, Baumeister et al. (2009) found that a similar manipulation was able to increase participants' aggression and decrease their helping behavior. These findings indicate that free will beliefs (FWBs) might be crucial for maintaining the motivation necessary to control selfish impulses in favor of pro-social behavior, in accordance with societal norms.

More recent work in the field of experimental psychology has revealed that the effects of weakening participants' FWBs are not restricted to complex social behavior, but even propagate to very basic levels of motor control (Lynn et al., 2013; Rigoni et al., 2011; 2013). Using a similar procedure to that of Vohs and Schooler (2008), Rigoni et al. (2011) found that inducing disbelief in free will was associated with a reduced amplitude of the readiness potential, an electrophysiological marker of pre-conscious movement preparation (Libet et al., 1983). In follow-up studies, it was found that weakening FWBs also influenced the effectiveness of other basic adaptive control processes, such as post-error slowing (Rigoni et al., 2013) or the intentional inhibition of pain avoidance behavior (Lynn et al., 2013). This suggests that weakening FWBs counteracts the recruitment of self-regulatory resources to adapt behavior in response to environmental demands.

However, despite these recent advances in research on the impact of FWBs on behavioral control, the mechanisms underlying the crosstalk between high-level beliefs

and low-level sensorimotor processes remain poorly understood. To explain their original finding regarding the readiness potential, Rigoni et al. (2011) speculated that weakening FWBs may reduce participants' *sense of agency* (SoA), i.e., the intrinsic experience of being in control of one's own actions (for reviews see Gallagher, 2000; Haggard & Chambon, 2012). This altered experience may then, in turn, hamper the recruitment of intentional effort for action production. Yet, despite the principle plausibility of this view, to date there exists only preliminary correlational supporting evidence.

Building on the notion that belief in free will often co-occurs with the pursuit of goal-directed behavior, Aarts and van den Bos (2011) tested the possibility that FWBs are associated with implicit processing of action-outcome relations underlying goal-directed behavior. The authors compared participants with either strong or weak dispositional FWBs in two different tasks that tapped into implicit aspects of agentic experience: (1) an intentional binding task, which measures the perceptual attraction of an intentional action and its sensory outcomes in terms of time, and (2) an action-outcome priming task, which assesses agency inferences resulting from a match between primed and actual outcomes (see Moore & Obhi, 2012; van der Weiden, Aarts, & Ruijs, 2013; for a review). Aarts and van den Bos (2011) found that strong dispositional FWBs were associated with greater intentional binding and a stronger influence of primes on agency inferences. These findings clearly indicate that FWBs and agency are related. In particular, they suggest that FWBs modulate the strength of predictive signals about action outcomes. Nevertheless, given the correlational nature of this study, a causal link between the two concepts remains to be established. Accordingly, the primary goal of the present study was to scrutinize the hypothesis that FWBs have a direct and causal impact on the SoA. Such evidence would provide the missing link to explain previous findings employing anti-free will manipulations, and highlight a general mechanism through which beliefs can affect even basic and unconscious adaptive processes. To this end, we employed the same procedure to manipulate the strength of FWBs as used in previous studies (e.g., Vohs and Schooler, 2008) in a within-subjects design. Participants were invited for two visits, in which they read essays that promoted either disbelief in free will or outlined general statements about consciousness (serving as a control condition).

Our secondary goal was to specify which aspects of agency are related to FWBs. It has been argued that the SoA constitutes a multi-dimensional construct comprising both an *implicit*, pre-reflective, or non-conceptual component that is related to lower-level perceptual and motor experiences, and an *explicit*, reflective or conceptual component that is related to higher-level thoughts and attributions (Synofzik et al., 2008). The findings by Aarts and van den Bos (2011) indicate that FWBs are related to implicit processes, yet, so far, it is unclear to what extent their influence may propagate to the conscious, deliberative level of explicit agency. Thus, in order to measure the effects of the induction procedure on participants' agency, we used two different experimental paradigms. To assess implicit components of agency, we employed an intentional binding task (see above and methods section). To assess explicit components of agency we used a modified version of the action-outcome learning method introduced by Sato and Yasuda (2005). This task requires participants to explicitly learn action-outcome relations and to subsequently rate their perceived agency over the outcomes. The benefit of this latter task is that it is explicit at all levels of processing; participants have full awareness of their actions and the corresponding agency cues (see methods section for a comprehensive description). Comparing the effects of inducing disbelief in free will in these two tasks should therefore reveal (i) if FWBs indeed have a causal impact on the sense of agency, and (ii) whether this causal link is directed towards implicit and/or explicit components of agency.

METHODS AND MATERIALS

Participants

Fifty-two students of Ghent University received a compensation of sixteen euros for their participation. The study was conducted in accordance with the Declaration of Helsinki, and the approval of Ghent University's Ethical Committee was obtained in advance. All participants reported being naïve as to the purpose of the experiment. One participant was excluded in advance of analysis for failing to return for their second session.

Procedure

Participants were individually tested during two sessions, taking place on the same weekday and time over two consecutive weeks. Each participant completed one '*control*' and one '*anti-free will*' (AFW) session, with session order counterbalanced across participants. Each session began with participants reading one of two possible excerpts from Francis Crick's *'The Astonishing Hypothesis.'* To ensure a thorough reading of the material and obfuscate the goal of the experiment, participants were informed that they would be tested regarding the material at two points during the session, and that memory retention was a major outcome of interest. In the control session, the excerpt was a brief historical overview of consciousness research, while the excerpt read during the anti-free will session questioned the reality of free will and posited that such a notion was pre-scientific (see Vohs and Schooler, 2008 for a similar procedure). Participants were invited to take as much time as they pleased to read and review the texts. Following the readings, participants were asked to complete the first of two tasks (either the IB task or the 'Sato' task, see below), the order of which was counterbalanced across participants. Subsequent to the first task, participants were asked to write a brief essay summarizing the previously read excerpt. Participants were then given the chance to reread the text prior to commencing the second task. After the completion of both tasks, participants were asked to fill out four questionnaires in their first session (the FAD+, the BIS/BAS, the LOC, and the PANAS-X) and two in the second session (the FAD+ and PANAS-X). The FAD+ (Paulhus & Carey, 2011) measures free will and deterministic beliefs, and served as the basis for our manipulation check. The PANAS-X (Watson, Clark & Tellegen, 1988) measures current mood. The BIS/BAS (Carver & White, 1994) is a measure of both behavioral inhibition and behavioral activation (cf. Gray, 1982), while the LOC (Rotter, 1966) measures a participant's locus of control. The latter questionnaires served exclusively as fillers to shield participants from discovering the purpose of our experiment following the first session. Once participants had completed the questionnaires they received a written quiz on the essay, and in their second session, a short debriefing. The quiz was not scored but only served to reinforce our cover story. The debriefing questionnaire consisted of general questions regarding the experiment: how participants felt about the duration of the

experiment, whether they felt their concentration slip, whether they had used any specific strategies, and finally, what they thought the purpose of the experiment was.

Manipulation check

To probe the general effectiveness of the manipulation in reducing participants' FWBs, we compared their scores on the free-will subscale of the FAD+ following the control and AFW sessions via paired samples *t*-tests (p = 0.05, one-tailed[2]). Moreover, we computed a session-based difference score on the free-will subscale as an individual screening of a given participant's response to the manipulation. In previous studies, we observed that a small but meaningful number of participants respond to the AFW manipulation in a reactant way, i.e., they reported a *stronger* belief in FW after reading an anti-free will text. Accordingly, we also wanted to explore how a reactant response affects the resulting level of agency. To this end, we conducted (explorative) post-hoc analyses in which the data of responders (i.e. participants whose FWBs were weakened in the AFW session) and reactant participants (i.e. participants whose FWBs were augmented in the AFW session) were analyzed separately.

Intentional Binding Task

As a measure of implicit agency components, we used a modified version of the 'intentional binding' method introduced by Haggard, Aschersleben et al. (2002) and Haggard, Clark et al. (2002). IB refers to the temporal attraction of an action and its sensory effects within the actor's perception (see Moore & Obhi, 2012, for a review). In this task, participants made time judgments about either actions or sensory events (tones) while these events occurred together or in isolation. They attended to the image of a centrally presented circular clock face (diameter = 8 cm) consisting of 60 dots (diameter = 2 mm). On every trial, a circular clock hand (diameter = 4 mm) rotated clockwise along the dots at a rate of 3 s per rotation, starting from an unpredictable clock hand position. Four different block types were employed, differing with respect to

[2] We employed one-tailed significance tests in all comparisons for which we had clear directional expectations. This is the case for the comparison of IB scores or agency ratings between the AFW session and CTR session (where we expected the AFW manipulation to decrease participants' agency). However, this is not the case for undirected comparisons (e.g., the test between the frequency of left- and right-hand responses in the acquisition phase of the Sato task).

the event to be judged (action vs. tone) and whether or not there was an instrumental relation between actions and tones (agency vs. baseline). In the agency conditions, participants were instructed to press a response key (keyboard space bar) with their right hand at a moment of their choosing. Their responses were followed by a brief sine wave tone (frequency = 600 Hz, duration = 75 ms) presented via headphones at a delay of 250 ms, while the clock hand continued rotating for an unpredictable interval (varied between 1000 and 2000 ms in steps of 250 ms) and then disappeared. After the clock hand disappeared, participants were prompted to indicate the perceived time of either their button press (agency action) or the tone onset (agency tone) by manually selecting (mouse clicks) the corresponding clock hand position. In the two baseline conditions, temporal judgments were made about actions (baseline action) or about tones (baseline tone) when these events occurred in isolation. Participants performed 20 trials for each block type, resulting in 80 trials overall. The block order was counterbalanced across subjects. To estimate individual IB scores, we first calculated the mean judgment errors (i.e., the deviance between perceived vs. actual time points of either actions or tones) for each block type and experimental session. Thereafter, binding scores were calculated for actions and for tones by subtracting judgment errors in the agency blocks from those in the corresponding baseline blocks. Finally, the overall amount of IB was computed by adding the absolute values of both binding scores (see Fig. 1 for a graphical illustration of the task). Mean judgment errors were analyzed in a general linear model (GLM) with the within-subjects factors JUDGMENT (action vs. tone), AGENCY (baseline vs. agency) and BELIEF CONDITION (control vs. AFW). To test whether IB was modulated by the FW manipulation, we conducted post-hoc paired-samples *t*-tests between the overall IB scores in the control session and the AFW session (one-tailed; $p < 0.05$).

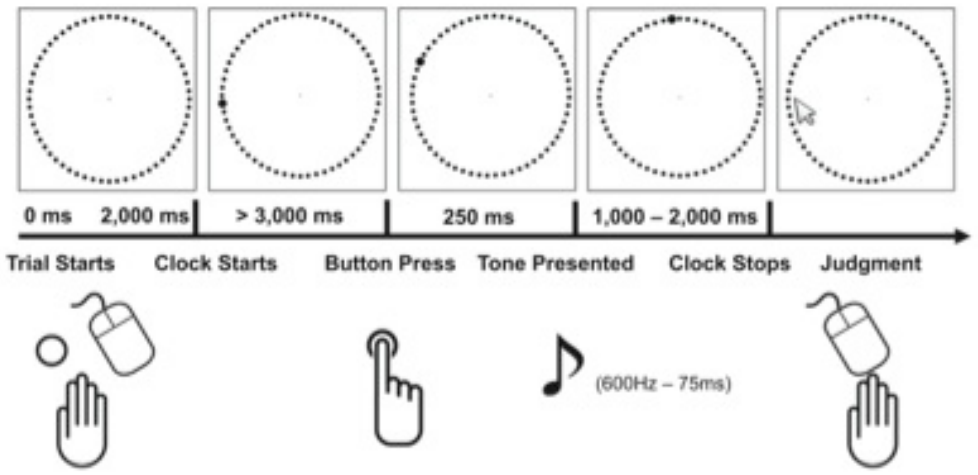

Figure 1. ***Illustration of the single-trial structure of the IB task (adapted from Demanet & Muhle-Karbe et al., 2013).***

Sato Task

To assess explicit components of agency, we employed an adapted version of the task introduced by Sato and Yasuda (2005). This rationale of this task is grounded in ideomotor theory (e.g., Prinz, 1997; Elsner & Hommel, 2001) and entails two distinct experimental phases. In an initial *acquisition phase*, novel action-effect associations are explicitly formed by means of frequent and contingent pairing of simple responses and resulting sensory outcomes. In a subsequent *test phase*, participants perform the same task and judge their perceived agency over the action outcome, while both the congruency of the action effect (compared to the acquisition phase) and its delay are manipulated.

In the acquisition phase, trials started with the presentation of the outline of a white square in the center of the screen. Participants were instructed to press either the NUM-del (with their right index finger) or NUM-enter button (with their right middle finger). Immediately after a response was given, the white square was replaced by a colored square with a specific color being assigned to each response (red and blue in one session, and yellow and green in the other; sequence of color sets counterbalanced across subjects). Participants were instructed to freely choose their responses in each trial, but to try to achieve an equal overall frequency of both responses without using a particular strategy such as simple alternations between left and right responses. Altogether, the acquisition phase comprised 200 trials, a number

of repetitions that has been shown to be sufficient to establish strong action-effect representations (see Elsner & Hommel, 2001).

In the test phase, participants were told that on some trials their action would cause the colored square to appear on the screen, while on other trials the computer would cause the colored square, and their task would be to infer the originator. Trials were similar to the acquisition phase with the following modifications: The color of the produced square was either congruent (i.e., the same color that a particular response produced in the acquisition phase) or incongruent (i.e., the color that was previously produced by the other response). Moreover, the delay between action and outcome was manipulated so that the colored square appeared on the screen either immediately after the response was given or at a delay of either 300 ms, or of 600 ms. Finally, after the colored square disappeared from the screen, participants were asked to rate their perceived agency over the action outcome (i.e., their certainty that they had caused the square to appear on the screen). To this end, a Likert scale was presented and participants indicated their answers on a scale from one to four, with one representing absolute certainty that the computer had produced the square and four representing absolute certainty that the square was produced by oneself. Ratings were given with the left hand fingers using the buttons 'Z', 'X', 'C', and 'V' of a QWERTY keyboard. Importantly, to avoid a contamination of the ratings by response biases, two different rating scales were presented across trials (randomly intermixed). These two scales were of opposite polarity (i.e., starting from 1 on the left to 4 on the right or vice versa). Prior to starting the test phase, participants were familiarized with the general task and the rating procedure. They first performed ten practice trials in which only the rating scales were presented followed by another ten practice trials with the complete task, supervised by the experimenter.

For both experimental phases, we first computed the proportion of left and right hand responses and the mean response times (RTs). Both scores were compared between the control session and the anti-free will session via paired-samples *t*-tests (two-tailed). Agency ratings of the test phase were analyzed in a GLM using the within-subjects factors CONGRUENCY (congruent vs. incongruent), DELAY (0 ms vs. 300 ms vs. 600 ms), and BELIEF CONDITION (control vs. AFW).

RESULTS

Belief manipulation

The comparison of participants' FAD+ scores between the control session and the AFW session confirmed that participants reported stronger determinist beliefs in the AFW session than in the control session, $t_{50} = 2.885$, $p = 0.003$ (AFW: $M = 3.01$, $SD = 0.59$; Control: $M = 2.88$, $SD = 0.64$). Moreover, the aforementioned screening procedure of individual difference scores (see methods section) identified eleven reactant participants who reported stronger FWBs in the AFW session than in the control session.

Intentional binding task

The GLM revealed a significant main effect of AGENCY, $F_{1,50} = 6.28$, $p = 0.015$, and a marginally significant trend of JUDGMENT, $F_{1,50} = 3.032$; $p = 0.088$. In addition, the interaction between JUDGMENT and AGENCY was significant, $F_{1,50} = 36.874$, $p < 0.001$, indicating that actions were perceived as occurring later in the agency blocks than in the baseline blocks (Judgment errors: Baseline $M = -0.184$ ms, $SE = 6.899$; Agency $M = 25.701$ ms, $SE = 8.578$), whereas the opposite was observed with tone judgments (Baseline $M = 58.217$ ms, $SE = 7.892$; Agency $M = -4.448$ ms, $SE = 11.922$). This pattern reflects a replication of the general intentional binding effect (Haggard et al., 2002). Most importantly, the three-way interaction between JUDGEMENT, AGENCY, and BELIEF was at trend-level, $F_{1,50} = 2.863$, $p = 0.097$, and reached significance in a one-tailed *t*-test between the IB scores of the control session (M = 100.116 ms SE = 18.104) and the AFW session (M = 76.973 ms, SE = 13.822), $t_{50} = 1.692$; $p = 0.048$. This one-tailed test can be justified by the directional expectation that weakening FWBs would diminish the sense of agency (see Fig. 2). All other main effects and interaction terms were non-significant (all *p*-values > 0.39).

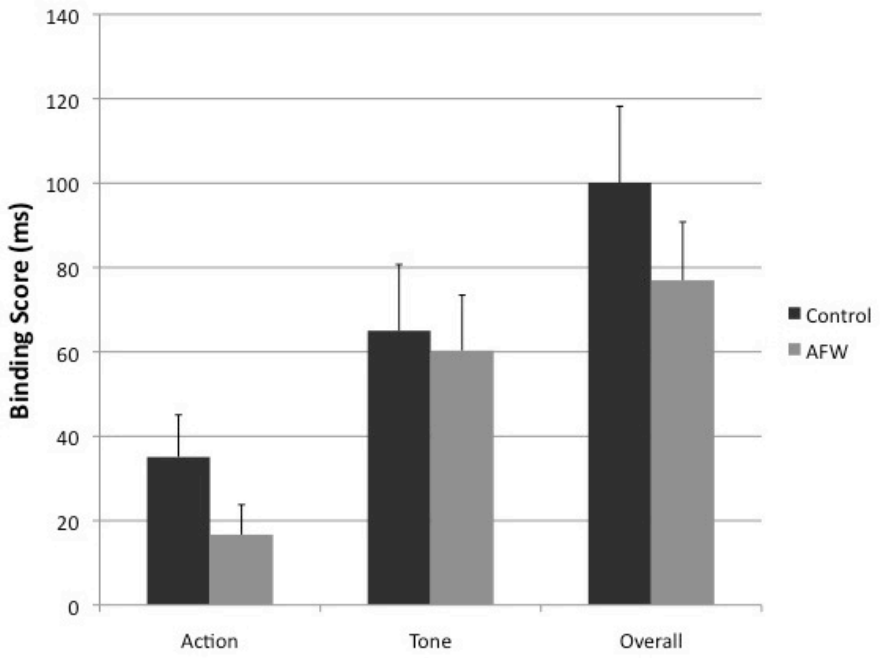

Figure 2. ***Results of the IB task. Bars display mean binding scores and standard errors for actions, tones and composite scores of both, separately for control and AFW conditions.***

Thereafter, we conducted explorative post-hoc analyses in which IB scores were analyzed separately for responders and reactant participants (see methods section). In the group of responders, the IB scores were significantly decreased in the AFW session (Control session M = 106.866 ms, SE = 21.202; AFW session M = 71.360 ms, SE = 15.065), t_{39} = 2.260; p = .015. By contrast, in the reactant group, IB scores did not differ and were in fact numerically reversed, with the AFW session showing greater binding scores (Control session M = 75.569 ms, SE = 33.856; AFW session M = 97.381 ms, SE = 34.015), t_{10} = .909; p =.193 (See Fig. 3). Thus, although only of exploratory value, this analysis indicates that the overall effect of the belief manipulation on IB was related to the individual effectiveness of the belief manipulation.

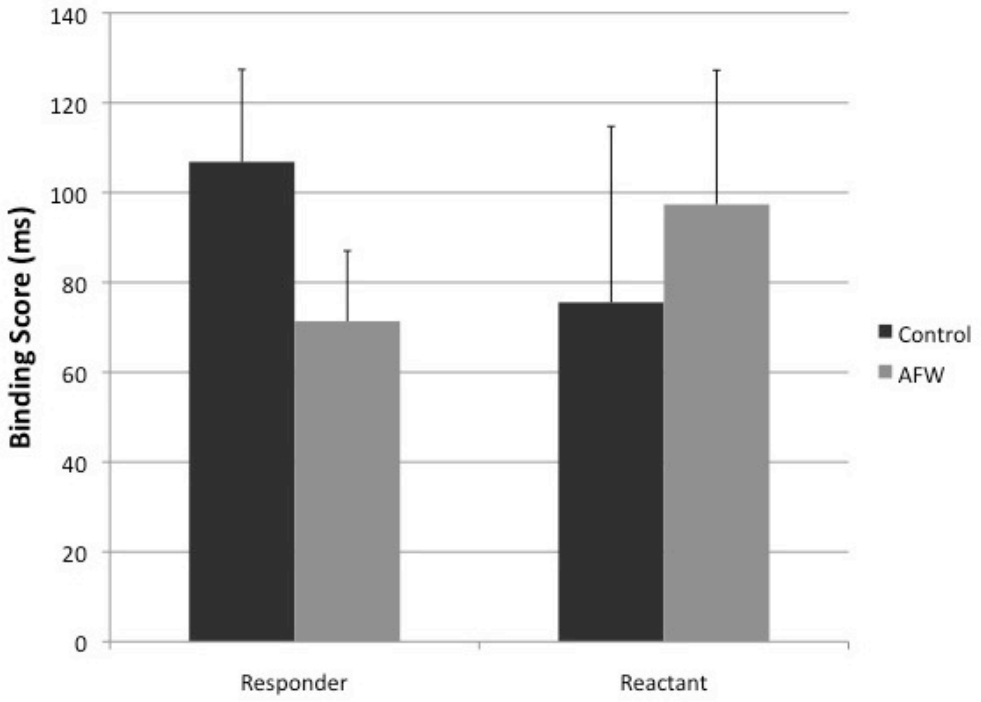

Figure 3. ***IB composite scores for reactant and responder participants, separately for control and AFW sessions. Error bars represent standard errors.***

Sato task: acquisition phase

The analysis of participants' RTs and response choices revealed that performance in the acquisition phase did not differ between the control session and the AFW session (RT: Control = 293.2 ms; AFW = 299.9 ms; percentage of left and right responses: Control = 49.6% vs. 50.4%; AFW = 51.8% vs. 48.2%; all *p*-values > 0.05).

Sato task: test phase

As for the acquisition phase, the proportion of response choices and the RTs did not differ between the two sessions (RT: Control = 320.1 ms; AFW = 326.6 ms; percentage of left and right responses: Control = 52.8% vs. 47.2%; AFW = 50.4% vs. 49.6%; all *p*-values > 0.05). Moreover, the analysis of participants' agency ratings revealed a significant main effect of CONGRUENCY, $F_{1,50} = 96.692$; $p < 0.001$, indicating higher agency ratings on congruent trials than on incongruent trials (Congruent *M* = 3.070 *SE* = .081; Incongruent *M* = 1.698, *SE* = .080). The main effect of DELAY was significant as well, $F_{1,50} = 48.022$; $p < 0.001$, reflecting a progressive decrease of agency rating with the length of the delay (0 ms delay *M* = 2.900, *SE* = .069; 300 ms delay *M* =

2.204, *SE* = .067; 600 ms delay *M* = 2.049, *SE* = .065). Both effects are replications of previous studies on explicit agency components (e.g., Sato & Yasuda, 2005; Spengler et al., 2009). In addition, there was a significant interaction between CONGRUENCY and DELAY, $F_{1,50}$ = 16.605; $p < 0.001$, reflecting a stronger delay effect for congruent effects than for incongruent effects (see Fig. 4). Importantly, neither the main effect of BELIEF was significant, $F_{1,50}$ = 1.118; p = 0.295, nor any interaction term involving this factor (all p-values > 0.203). Thus, our results replicated previous observations that congruency with prior action-effect contingencies, along with as the delay of the action outcome, affects the explicit components of agency (Sato & Yasuda, 2005; Spengler et al., 2009). By contrast, agency ratings were not affected by the manipulation of FWBs.

As in the IB section, we next conducted the post-hoc analyses in which the data of responders and reactants were analyzed separately. In both groups, however, the same pattern was evident as in the main analysis (i.e., significant main effects of CONGRUENCY, DELAY, and a significant interaction between the two factors, but non-significant main effects of BELIEF and interaction terms involving this factor). Thus, contrary to the IB data, explicit agency ratings were not influenced by the effectiveness of the belief manipulation.

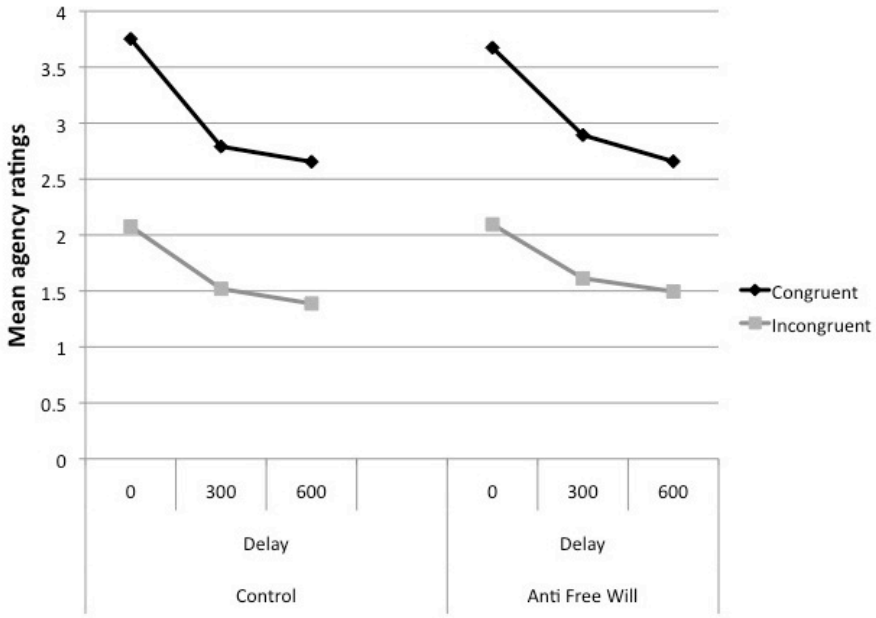

Figure 4. ***Results of the Sato task. Values indicate mean agency ratings as a function of congruency, delay, and belief conditions.***

DISCUSSION

The aim of the present study was to probe whether weakening belief in the concept of free will would have a causal impact on participants' sense of agency. In line with our hypothesis, we found that IB was significantly reduced in the AFW condition, indicating that determinist beliefs hamper the implicit sensation of being in control of one's actions. By contrast, participants' explicit agency ratings were not affected by the belief manipulation.

Free will beliefs and self-control

The present study complements and extends a growing body of research on the interaction between high-level determinist beliefs and low-level processes involved in self-control. Previous studies have shown that participants induced to disbelieve in FW exhibit less involvement in motor preparation (Rigoni et al., 2011), do not adapt their behavior in response to unwanted action outcomes (Rigoni et al., 2013), and are less inclined to engage in effortful cancellation of prepotent behavior (Lynn et al., 2013; Rigoni et al., 2012). Our findings add several new and valuable insights to this existing literature. First and foremost, establishing a causal link between determinist beliefs and the sense of agency provides an integrative mechanism that explains how FWBs impact behavioral control in a variety of contexts. It is possible that FWBs are able to influence the entire action cycle by intervening at an early or pre-reflective level. This impact on sensorimotor binding then cascades to more overt behavior, eliciting in turn less intentional effort, a reduced sense of agency, and less feeling of responsibility. Interestingly, a recent study has linked the IB effect to the feeling of responsibility (Moretto et al., 2011). These authors employed a modified IB paradigm in which actions had unpredictable consequences, either in a moral context or a simple economic context. IB was enhanced when actions were embedded in a moral context, suggesting binding is sensitive to a feeling of responsibility over self-produced action consequences. It is tempting to speculate that an intrinsic bias to bind actions with their outcomes in time could constitute a building block of higher-order social cognition, and in light of these findings it becomes more apparent why the manipulation of FWBs can impact such a wide range of behaviors.

Finally, from a methodological perspective, our study is the first to employ a successful within-subjects manipulation of FWBs. As such, it provides more direct evidence for the causal impact of high-level beliefs, since pre-existing differences between different groups of participants can be ruled out as an alternative explanation. Moreover, having established the feasibility of this manipulation in within-subjects designs may permit its application in new experimental settings, e.g., in combination with brain imaging techniques such as fMRI.

Implicit and explicit components of agency

Beyond establishing a causal link between FWBs and agency, our data revealed striking differences between implicit and explicit components of agency. While IB scores were reduced in the AFW session, the explicit agency ratings in the Sato task were not affected. This dissociation is interesting for several reasons. First, it indicates that the influence of FWBs on agency does not extend to the level of reflective deliberation. Instead, beliefs seem to bias very basic and implicit processes that underlie our pre-reflective self-perception as intentional agents. In cases where the integration of agency cues into judgment is transparent, explicit processes may override the modulation of pre-reflective prediction processes. In addition, the dissociation between implicit and explicit agency components also speaks against the possibility that our data may originate from "demand effects" (i.e., participants form expectations about the purpose of the experiment and try to please the experimenter by fulfilling his/her hypotheses). Finally, on a theoretical level, this result corroborates the notion that the sense of agency is not a unitary psychological concept, but rather entails distinct components that tap into different facets of action awareness (see David et al., 2008; Obhi & Hall, 2011; Synofzik et al., 2008). In line with this view, a recent study indicated that explicit and implicit components might operate at different time scales. Ebert and Wegner (2010) employed a paradigm that simultaneously assessed implicit and explicit agency components in a relatively naturalistic task setting. Participants were instructed to push or pull a lever, causing objects on a screen to either come closer or move further away from them. As in the present study, the authors manipulated the congruency between actions and outcomes, and the outcome delay. It was found that both measures were influenced by action-outcome congruency (i.e., both implicit and explicit agency was

higher for congruent outcomes). However, the influence of congruency was more pronounced for explicit agency, particularly with long outcome delays, suggesting that implicit components may operate on a shorter time scale.

Limitations and outlook

Despite the clear overall patterns of results, several aspects of our data indicate the need for further investigation. First, it must be considered that the effect of the belief manipulation on IB reached significance only in a one-tailed *t*-test. The post-hoc analyses indicated that this was likely due to the variability in participants' responses to the manipulation: There was a robust attenuation of IB in the group of responders, but in the group of reactant participants this effect was not only absent but even reversed numerically. On the one hand, this finding is reassuring, as it confirms that the effect is related to the effectiveness of the belief manipulation. On the other hand, it raises new questions, in particular what determines an individual's response, and whether a reactant response only eliminates the effects of the FW manipulation or if it can even increase agency as compared to a control session. The rather small number of reactant subjects in our sample prohibits us from making conclusions about this issue. Furthermore, the precise nature of the implicit processes responding to FWBs remains to be uncovered through future research. While IB has primarily been related to motor prediction, the findings of Aarts and van den Bos (2011) indicate that non-motor predictive and inferential processes might be affected as well.

Conclusion

The present study revealed that inducing disbelief in the concept of FW has a causal impact on implicit components of agency, but not on explicit agency components. This finding points to a psychological mechanism through which determinist beliefs exert their wide-ranging influence on human behavior, and highlights the multi-dimensional nature of the sense of agency.

Acknowledgements

This work was supported by the European Science Foundation's EUROVETO project (09-ECRP-020).

REFERENCES

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: enhancing experienced agency by priming effect information. *Consciousness and Cognition*, *14*, 439–458.

Aarts, H., & van den Bos, K. (2011). On the foundations of beliefs in free will: intentional binding and unconscious priming in self-agency. *Psychological Science*, *22*, 532–537.

Baumeister, R. F., Masicampo, E. J., & Dewall, C. N. (2009). Prosocial benefits of feeling free: disbelief in free will increases aggression and reduces helpfulness. *Personality & Social Psychology Bulletin*, *35*, 260–258.

Carver, C. S. & White T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333.

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition, 17*, 523-534.

Demanet, J., Muhle-Karbe, P. S, Lynn, M. T., Blotenberg, I., & Brass, M. (2013). Power to the will: How exerting physical effort boosts the sense of agency. *Cognition, 129*, 574-578.

Ebert, J. P. & Wegner, D. M. (2010). Time warp: authorship shapes the perceived timing of actions and events. *Consciousness and Cognition, 19*, 481-489.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology. Human Perception and Performance*, *27*, 229–240.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*, 14-21.

Gray, J. A. (1982). *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system.* New York: Oxford University Press.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*, 934–946.

Haggard, P., Aschersleben, G., Gehrke, J., & Prinz, W. (2002). Action, binding, and awareness. *Common mechanisms in perception and action: Attention and Performance XIX,* 266-285.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*, 390-392.

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

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, 106*, 623–642.

Lynn, M. T., Van Dessel, P., & Brass, M. (2013). The influence of high-level beliefs on self-regulatory engagement: evidence from thermal pain stimulation. *Frontiers in Psychology*, *4*, 614.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition*, *21*, 546–561.

Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. *Consciousness and Cognition*, *20*, 1847–1854.

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment*, *93*, 96–104.

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology, 9*. 129-154.

Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition*, *21*, 1482–1490.

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*, *22*, 613–618.

Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition*, *127*, 264–269.

Rotter, J. B. (1966). Generalized expectancies for internal versus external locus of control reinforcement. *Psychological Monographs: General and Applied, 80.* Whole No. 609.

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94*, 241–255.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J.-D. (2008). Unconscious deteminants of free decisions in the human brain. *Nature Neuroscience, 11*, 543-545.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *NeuroImage*, *46*, 290–298.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*, 219–239.

Van der Weiden, A., Aarts, H., & Ruys, K.I., (2013). On The Nature of Experiencing Self-Agency: The Role of Goals and Primes in Inferring Oneself as the Cause of Behavior. *Social Psychology and Personality Compass, 7,* 888-904.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49–54.

Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, *27*, 649–659.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAs scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

The goal of this dissertation was twofold. First, I wanted to develop an experimental paradigm that allows for the investigation of intentional inhibition in a more ecologically valid context. Furthermore, I wanted to investigate the determinants of volitional engagement. Below, I give a brief overview of the empirical results of each section of the thesis, followed by a discussion of limitations and possible directions for future research.

Impulse control

Although research on response inhibition has a long tradition in cognitive psychology and neuroscience, the functional and neural basis of impulse control (in realistic scenarios) is still insufficiently understood; in classical experimental paradigms, participants' behavior is completely determined by external stimuli (e.g., Aron & Poldrack, 2006; Chambers et al., 2006; Chikazoe et al., 2007; 2009). This rationale allows for the necessary and precise delineation of componential processes, but the transfer of findings to realistic instances of impulse control is rather questionable (see Haggard, 2008). To overcome this problem, the concept of intentional inhibition has been introduced, and corresponding paradigms require participants to make endogenous decisions about acting or inhibiting (Brass & Haggard, 2007; 2008). Preliminary studies from this young line of research clearly require internal decision-making (Brass & Haggard, 2007; Kühn et al., 2009; 2013; Walsh et al., 2010); however, decisions are often rather arbitrary and absent urges or a prepotency of action. Hence, the primary goal of the first part of my thesis was to develop a novel paradigm that permitted the measurement of behavioral inhibition in a more ecologically valid manner. This endeavor was largely successful; applying thermal pain stimulation in the MR scanner proved feasible, and the experimental parameters successfully induced a level of pain sufficient to provoke a substantial urge to act in each participant and throughout the whole experiment. Concurrently, the individually-tailored pain levels were moderate enough to allow for voluntary engagement in a choice scenario. Hence, the paradigm provided a useful and novel setting for the investigation of impulse control, which justifies the additional efforts that the use of thermal stimulation entails (e.g., the extensive piloting and pre-testing to determine individual pain thresholds).

Similarities between externally-guided and intentional inhibition

One of the central observations resulting from this new paradigm was that the difference between endogenous and exogenous inhibition decisions dissipated. Specifically, both types of inhibition relied on the joint activation of brain areas that were previously implicated in either externally-guided inhibition (see Aron et al., 2014) or in intentional inhibition (see Brass & Haggard, 2007; Kühn et al., 2009; 2013). Thus, rather than performing independent functions, these areas appeared to work in concert to enable suppression of the pain avoidance response. This finding is at odds with earlier proposals of two distinct neural circuitries for inhibition (Brass & Haggard, 2007; 2008) and indicates the need for their comprehensive integration.

In reconsidering previous work on the two types of behavioral inhibition, it appears that the respective studies have captured complementary components of inhibitory control, each of which may contribute to the control of impulses in ecologically valid situations. External inhibition studies focus on the implementation component of inhibition by requiring participants to cancel an already initiated action. This reflects the final stage of impulse control, namely the suppression of an impulsive act after the decision to inhibit has been determined (in this case by the external stop or nogo signal). By contrast, intentional inhibition studies focus more on the decisional components of inhibition by requiring participants to choose between engaging in an action and withholding from doing so. This reflects a much earlier stage of impulse control. Importantly, in realistic instances of inhibitory self-regulation both components play a critical role and must necessarily interact, particularly when a very strong urge must be suppressed continuously rather than instantaneously. In the current paradigm, thermal stimulation persisted for several seconds, likely necessitating a continuous reinforcement of both decisional and implementational aspects to adhere with the demand to endure the pain (see Ridderinkhof et al., 2014).

The distinction between early decisional and late implementational aspects of impulse control is consistent with the differing connections of the dFMC and the rIFG. Whereas the rIFG is well-suited to implement motor control due to its rich connections to the motor cortex (Aron et al., 2007; 2011), the dFMC has connections with diverse areas, including those implicated in processing internal, emotional, and semantic information (Bzdok et al., 2012; Eickhoff et al., 2014). An interesting prediction of this

dual-component view on inhibitory control is that the dFMC should be the source of signals that instantiate the implementation of inhibition and thus recruit the stopping system. Previous findings have yielded positive evidence that the dFMC is functionally coupled with the preSMA during intentional inhibition (PPI analysis; Kühn et al., 2009; 2013), but further research seems necessary to reveal the temporal and functional dynamics within the extended inhibition network.

The mechanisms underlying intentional inhibition

Although the dFMC has been reliably associated with intentional inhibition, the matter of how it supports this capacity remains controversial. Seminal studies on intentional inhibition suggested that activity of the dFMC directly reflects the suppression of motor output (*veto signals*). For instance, Brass & Haggard (2007) employed a variant of the Libet task in which participants chose between pressing and not pressing a button while monitoring the timing of their motor intentions. The authors observed not only increased dFMC activity on trials in which participants chose not to respond, but also a negative correlation between activity in the dFMC and activity in the primary motor cortex. To some extent, the data from the pain avoidance suppression paradigm support and extend this view of intentional inhibition as motor suppression; the ROI analysis in Chapter 1 that focused on the motor hand area could reveal a trace of inhibition (i.e., reduced activity on inhibition trials). Interestingly inhibition was directed only towards the effector hand (i.e., the hand that could terminate the pain indirectly via button press), but not towards the stimulated hand (which could theoretically also terminate the pain, via withdrawal). However, these findings need to be taken with caution, as the stimulated hand was restrained via a sandbag to prevent automatic withdrawal, and it remains to be tested if the same results would be found in an unconstrained setting. Moreover, although these findings clearly indicate that motor inhibition took place, the source of this inhibition remains elusive and could possibly be located outside the dFMC.

As discussed in Chapter 2, it is quite noteworthy that activation of the dFMC has been documented in a variety of other situations that entail the demand to suppress impulses or urges, but no actual motor behavior. For example, dFMC activation has been found with the endogenous suppression of negative emotions (Kühn et al., 2013),

cigarette cravings (Brody et al., 2007; Hanlon et al., 2013; Hartwell et al., 2012), or even gambling desires (Campbell-Meiklejohn et al., 2008). To reconcile the involvement of the dFMC in the suppression of motor behavior as well as more abstract desires and urges, two different solutions have been formulated. Filevich et al. (2012) proposed that even more abstract states such as emotions or desires might also be coded in terms of associated motor programs (e.g. facial muscles used to express a particular emotion). In this sense, inhibition of the corresponding urge may still have a motor target, albeit a more abstract target. Alternatively, Chapter 2 discusses the idea that the dFMC may serve a more general role in human self-control that reflects general disengagement rather than motor control in a strict sense (Lynn et al., 2014). More research will be necessary to evaluate both accounts.

Future studies should also consider the possibility of different subdivisions within the large dFMC region. As noted above, the dFMC peak locations reported in previous intentional inhibition studies varied considerably in their exact location. This could reflect the spatial uncertainty of neuroimaging experiments, but it could also indicate the existence of different context-dependent control mechanisms. In line with the latter view, a recent meta-analytic parcellation study revealed four different subdivisions of the dFMC that could be isolated on the basis of their distinct functional connections (Eickhoff et al., 2014). Either way, future research should take these neuroanatomical landmarks into account in order to maximize spatial accuracy.

Limitations and future directions

In spite of the aforementioned strengths of this paradigm, several important challenges remain to be addressed with further research. For instance, a certain degree of ambiguity exists with regard to the timing of the choice between acting and inhibiting. We could not fully exclude the possibility that participants decide in advance of each trial how they will behave in the event of a choice cue, which would call into question whether the analysis of cue-locked brain activity captures the proper time period. In fact, this is a classical problem when relying on internally-generated decisions in experimental studies (see Nachev & Hussain, 2010; Nachev & Hacker, 2014). Although several aspects of our data indicate that participants in the study of Chapter 1 did not make their choice prior to the cue-onset (e.g., the increased RTs on choice

action trials compared with externally guided action trials), it would nonetheless be desirable to find a method that allows for more precise tracking of the decision process. Furthermore, it remains an open issue to what extent the results from the current paradigm are specific to the suppression of pain avoidance behavior or instead generalize to other domains of self-control. Although we put considerable efforts into the dissociation of pain-related from inhibition-related brain activity, more research will be necessary to evaluate the generality of the neural basis of self-control. To this end, studies will be needed that assess the suppression of other behavioral urges, and the challenge will lie in finding a way of inducing urges with similar reliability. The positive association between inhibition-related activity in the present paradigm and scores on the Self Regulation Questionnaire (observed in Chapter 3) is certainly encouraging in this respect, and indicates that our findings might generalize to other self-regulatory behaviors.

A final issue pertains to the type of action decisions in our paradigm. One may argue that the choices employed here still suffer from some of the shortcomings that we sought to address, albeit to a lesser extent. In particular, the choice between acting impulsively (to terminate pain) and withholding from executing this impulse differs in an important respect from many instances of realistic self-control, as there are no genuine (positive) incentives for inhibiting. Participants presumably occasionally opt to endure the pain only because of the desire to adhere to the experimental instructions to achieve a similar frequency of action and inhibition trials. However, in real life, self-control is often characterized by a conflict between different incentives for acting impulsively (e.g., the pleasure of a fun activity versus work) and for controlling these impulses (e.g., the prospect of advancing one's career). The incentives for self-control are often more abstract or temporally distal in nature than those of their corresponding impulses; however, these countering incentives do exist and can be considered crucial motivators for the exertion of effortful self-control. The circumstances of the present paradigm rather mirror situations in which impulses have to be controlled in favor of social norms (e.g. not stealing a desired but inaccessible object), where the benefit of self-control is the *absence* of a negative outcome rather than the accomplishment of a positive outcome. Accordingly, a very interesting but challenging perspective for future research would be to incorporate abstract positive incentive structures for acts of self-

control into experimental paradigms in order to mirror the portrayed self-regulatory scenarios more closely.

Trait and state levels of self-control

Chapter 3 used the pain avoidance paradigm to test whether inhibition-related brain activity is associated with individual differences in self-regulatory ability, both at the state level and at the trait level. Importantly, trait differences in self-regulation were positively associated with activity in several inhibition areas, i.e., they were recruited for inhibition to a greater extent by people with high self-regulatory abilities. This result is encouraging for several reasons. Methodologically, it indicates that the paradigm is indeed sensitive to aspects of self-regulation that are required in realistic scenarios. Theoretically, it indicates that individual differences in self-regulatory abilities might reflect the differential ability to recruit inhibitory brain regions. Future research should further expand on this, for example by investigating individual differences in more specific self-control strategies such as distraction and reappraisal (e.g., Metcalfe & Mischel, 1999). In contrast to this positive result, no effects of the ego depletion manipulation were found in Chapter 3, neither behaviorally nor at the brain level. This null effect is somewhat difficult to interpret, particularly in light of positive results in several behavioral studies employing similar depletion methods. One has to consider the possibility that the scanner environment counteracted the effectiveness of ego depletion, either via recruitment of additional resources (as a consequence of stress), via recovery (due to the temporal delay between depletion task and scanner task), or due to resource management on the part of participants in anticipation of the upcoming pain. Thus far, only one study has successfully demonstrated a depletion effect in a neuroimaging study (Wagner & Heatherton, 2012), although the outcome measure (amygdala activation during an emotional categorization task) is arguably only indirectly related to self-regulation; most ego depletion studies have instead directly investigated the impact of depletion on a subsequent task requiring self-control. It should also be noted that recent research has called the concept of ego depletion into question. Publication bias often prevents null results from being disseminated, but a few recent studies nevertheless suggest that the ego depletion effect is rather weak if not entirely absent (Carter & Mccullough, 2014; Xu et al., 2014). Thus, at present, the

phenomenon of ego depletion seems to lack the robustness to encourage further use in neuroimaging studies, at least until behavioral research has further specified its boundary conditions. Abundant evidence furthermore indicates that the depletion effect seems to rely on mediating factors such as an individual's beliefs, both about the veracity of the limited resource model and about their own level of depletion (Clarkson, Hirt, Jia, & Alexander, 2010; Job, Dweck, & Walton, 2010). Beliefs seem therefore to have a nontrivial impact on self-regulation and volition, and with that in mind, in Chapters 4 and 5 we directly manipulate participants' belief structures.

Determinist Beliefs

Repeated experiments have shown that determinist beliefs negatively impact behavior (e.g., eliciting increased cheating and aggression; Vohs & Schooler, 2008; Baumeister et al., 2009). These findings also extend to low-level sensorimotor processes (Rigoni et al., 2011) and engagement in voluntary inhibition (Rigoni et al., 2012). Presumably, weakening free will beliefs diminishes the recruitment of cognitive resources for self-regulation at an early stage of processing. In Chapter 4, we investigated whether this explanation might apply to effortful impulse control in a motivational scenario requiring a high level of engagement. Using a pain paradigm similar to those of Chapters 1 and 3, we manipulated free will beliefs on a trial-by-trial basis in a between-subjects design.

In line with our predictions, participants who were induced to disbelieve in free will were significantly slower to terminate the pain stimulation on choice action trials, while there was no difference between groups on externally-instructed action trials. This suggests a specific impairment in voluntary action selection for self-regulation, rather than a global passivity on the part of participants. On an individual level, the amount of slowing on choice trials was furthermore correlated with the effectiveness of the belief manipulation, mirroring the finding by Rigoni et al. (2011) in which decreases in the readiness potential were correlated with a change in anti-free will scores. Anti-free will participants furthermore reported greater urges to terminate the trial when their behavior was guided by the cue compared to when they were able to freely choose, suggesting a disengagement from the task when externally instructed. Importantly, these differences were not confounded by differential response biases, as

the proportion of inhibition in choice trials was equivalent between groups. Contrary to our expectations, participants did not appear to use recent pain as a criterion in deciding whether to act or inhibit on choice trials. However, the experimental groups did differ in terms of their response styles; anti-free will participants tended to inhibit more often following an action trial than an inhibition trial, while this was not the case for control participants. Interpretations are merely speculative without further research; the choice behavior of anti-free will participants may reflect a more explicit desire to conform to the 50% instruction, or control participants may be selecting a more adaptive strategy by choosing the subjectively less effortful response (response repetitions vs alternations). This may be disentangled in future studies by presenting blocks composed solely of choice trials in order to determine, via longer choice trial sequences, which is the favored strategy.

Overall, the results of Chapter 4 suggest that the reduction of free will beliefs corresponds with a reduction in effort investment that influences voluntary action selection and inhibition, and a disinclination to adapt one's behavior to contextual needs. A fundamental belief in control over one's actions may therefore prove to be an integral prerequisite for self-regulatory investments. Yet knowing the impact beliefs have on self-control was only a piece of the puzzle; the means by which this effect is achieved remained poorly understood.

With that in mind, in Chapter 5 we sought to directly address the underlying mechanism by which beliefs impact the recruitment of self-control. To explain their original finding regarding the readiness potential, Rigoni et al. (2011) speculated that weakening free will beliefs may reduce participants' *sense of agency* (SoA), i.e., the intrinsic experience of being in control of one's own actions. This altered experience may then impede the recruitment of intentional effort for action production. In the first study to investigate the relationship between free will beliefs and agency, Aarts and van den Bos (2011) found that strong dispositional free will beliefs were associated with greater intentional binding and a stronger influence of primes on agency inferences, suggesting that free will beliefs modulate the strength of predictive signals about action outcomes. Nevertheless, given the correlational nature of this study, a causal link between the two concepts remained to be established. To that end, in Chapter 5, we tested the hypothesis that free will beliefs have a direct and causal impact on the sense

of agency, using a fully within-subjects design. Our secondary goal was to specify which aspects of agency are related to free will beliefs; Synofzik et al. (2008) proposed that the sense of agency consists of an implicit, pre-reflective component and an explicit, conceptual component. The findings of Aarts and van den Bos (2011) demonstrated that FWBs are related to implicit processes, yet it was unclear to what extent their influence might propagate to the conscious, deliberative level of explicit agency. Thus, we used two tasks to measure the effects of the induction procedure, each intended to tap into a different component of agency.

We found that intentional binding (the implicit task) was significantly reduced in the anti-free will condition, indicating that determinist beliefs hamper the implicit sensation of being in control of one's actions. By contrast, participants' explicit agency ratings were unaffected by the belief manipulation. Establishing such a causal link between determinist beliefs and the sense of agency points to an integrative mechanism by which beliefs are able to influence the entire action cycle by intervening at an early stage of processing before cascading to more overt behavior, including error-related behavioral adjustments (feedback processing; Rigoni et al., 2013; Rigoni, Pourtois, & Brass, 2014). It is possible, then, that the intrinsic bias to temporally bind actions with their outcomes constitutes a key component of higher-order social cognition.

Limitations

Chapters 4 and 5 both clearly illustrate that the effects of free will manipulations are far from universal and depend strongly on individual subject variables. In Chapter 4, the effects were only present in those participants who reported a level of pain sufficient to require self-control to execute inhibition. In Chapter 5, the effect was absent (numerically even reversed) in a subgroup of participants who responded in a reactant way to the manipulation. These observations illustrate that future research should shift from merely identifying novel behaviors that are linked to the strength of beliefs towards specifying the mechanisms that determine the effects of the manipulation at a single-subject level. Presumably this will necessitate taking into account participants' prior belief system, their mood, and their expectations with regard to outcomes.

Conclusion

In summary, the volitional components under investigation in this thesis have been elucidated in several ways. Behavioral inhibition under ecologically valid circumstances seems to elicit a form of inhibition that recruits both stopping networks, suggesting that intentional and externally-guided inhibition are not dichotomous, but rather exist along a continuum. Second, a proposal for the putative role of the dFMC in inhibition has been generated in light of this and related findings. This proposal posits that the dFMC serves general disengagement from urges and impulses to facilitate successful self-regulation. Third, the magnitude of inhibition-related activity in core regions of the extended inhibition network seems to be associated with trait differences in self-regulatory abilities, yielding a promising avenue for future research.

Finally, the findings of Chapters 4 and 5 support and extend previous research on free will beliefs and basic cognitive processes by revealing that weakening the strength of beliefs reduces intentional engagement in self-control and attenuates implicit components of agency. These results also complement the proposal that free will beliefs have an adaptive function in controlling human behavior, and undermining their strength reduces volitional engagement. The findings reported in Chapter 5 are particularly informative, as the established link between free will beliefs and agency might provide a missing link to explain previous findings; a reduced intrinsic bias to bind actions with their outcomes could represent the first step in a cascade of events subsequent to free will manipulations, such as diminishing the investment of intentional effort in action production, the responsiveness to environmental feedback, and the ability to control behavioral impulses.

REFERENCES

Aarts, H., & van den Bos, K. (2011). On the foundations of beliefs in free will: intentional binding and unconscious priming in self-agency. *Psychological Science*, *22*, 532–537.

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biological Psychiatry, 69*, 55-68.

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusing-weighted magnetic resonance imaging. *Journal of Neuroscience, 27*, 3743-3752.

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*, 2424-2433.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences, 18,* 177-185.

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Pro-social benefits of feeling free: Manipulating disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin, 35,* 260-268.

Brass, M. & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *Journal of Neuroscience*, *27*, 9141-9145.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuoscientist, 14,* 319-325.

Bzdok, D., Langer, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience,* 7: 232, 1-17.

Carter, E. C. & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology,* 5: 823, 1–11.

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., Morris, A. P., & Mattingley, J. N. (2006). Executive “brake failure” following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience, 18,* 444-455.

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activatin of the right inferior frontal gyrus during response inhibition across response modalities. *Journal of Cognitive Neuroscience, 19,* 1-12.

Chikazoe, J, Jimura, K., Asari, T., Yamashita, K., Morimoto, H., Hirose, S., Miyashita, Y., & Konishi, S. (2009). Dunctional dissociation in right inferior frontal cortex during performance of go/no-go task. *Cerebral Cortex, 19*, 146-152.

Clarkson, J. J., Hirt, E. R., Jia, L., & Alexander, M. B. (2010). When perception is more than reality: the effects of perceived versus actual resource depletion on self-regulatory behavior. *Journal of Personality and Social Psychology, 98*(1), 29 – 46.

Eickhoff, S. B., Laird, A. R, Fox, P. T., Bzdok, D., & Hensel, L. (in press). Functional segregation of the human dorsomedial prefrontal cortex. *Cerebral Cortex.*

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience, 9*, 934-946.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion – Is it all in your head?: Implicit theories about willpower affect self-regulation. *Psychological Science, 21*(11), 1686 – 93.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: how the "veto area" exerts control. *Human Brain Mapping, 30,* 2834-2843.

Kühn, S., Haggard, P., & Brass, M. (2013). Differences between endogenous and exogenous emotion inhibition in the human brain. *Brain Structure and Function, 219*, 1129-1138.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106,* 3 – 19.

Nachev, P., & Hacker, P. (in press). The neural antecedents to voluntary action: a conceptual analysis. *Cognitive Neuroscience.*

Ridderinkhof, K. R., van den Wildenberg, W. P. M., & Brass, M. (in press). "Don't" versus "won't": principles, mechanisms, and intention in action inhibition. *Neuropsychologia.*

Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition, 21*(3), 1482–1490.

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, 22,* 613-618.

Rigoni, D., Pourtois, G., & Brass, M. (2014). "Why should I care?" Challenging free will attenuates neural reaction to errors*. Social Cognitive and Affective Neuroscience,* nsu068.

Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition, 127,* 264-269.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*, 219–239.

Vohs, K. D., & Schooler, J. (2008). The value of believing in free will: Encouraging a belief in Determinism increases cheating. *Psychological Science, 19,* 49–54.

Wagner, D. D. & Heatherton, T. F. (2012). Self-regulatory depletion increases emotional reactivity in the amygdala. *Social Cognitive and Affective Neuroscience,* nss082.

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*, 619-626.

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., Middleton, K. R., Wing, R. R. (2014). Failure to replicate depletion of self-control. *PLoS ONE, 9*(10), e109950.

Wil: Van zelf-controle tot 'agency'

Waaruit bestaat het gevoel van wil (nl., het vermogen van keuze en determinatie waarmee we onszelf of onze omgeving controleren), met aanduidingen als *willen*, *beogen, wensen,* en *geneigd zijn om*? Het is het subtiele verschil tussen een oogknippering en een knipoog (Morsella, Molapour, & Lynn, 2013). De wil is zonder twijfel fundamenteel aan wat het betekent om mens te zijn; we verschillen van andere soorten omwille van dit specifieke vermogen om dingen in overweging te nemen. Terwijl het gedrag van andere soorten grotendeels bepaald wordt door omgevingstriggers, bevrijdt onze mogelijkheid om gedrag te controleren ons van onmiddellijkheid (Shadlen & Gold, 2004).

Problemen inherent aan wil als wetenschappelijk concept

Door het gebrek aan een alwetend inzicht in de interne toestand van anderen zijn wetenschappers geconfronteerd met de worsteling van het definiëren wat een wilsdaad is en hoe dit gemeten kan worden. Fenomenologisch is een wilsdaad gekarakteriseerd door de eerste-persoonsontologie: Ik doe iets uit eigen wil als ik voel dat ik gehandeld heb zoals ik het bedoelde. Deze conceptie van wil wordt gereflecteerd in de introspectieve traditie van de vroege dagen van empirisch psychologieonderzoek (e.g., Harleß, 1861; James, 1890; Lotze, 1852). Echter, sinds de opkomst van het behaviorisme werd introspectie afgeschreven omwille van de subjectiviteit en grotendeels verworpen als methode voor psychologisch onderzoek. Vrijwillig gedrag kan aan de andere kant ook gedefinieerd worden door het te vergelijken met gedrag dat uitgelokt wordt door een stimulus (e.g., Herwig, Prinz, & Waszak, 2007; Herwig & Waszak, 2009; Waszak et al., 2005). Deze definitie is vruchtbaarder voor empirisch onderzoek, gezien het toelaat acties te vergelijken die verschillen in de locus van de beslissing, terwijl ze toch gelijkaardig zijn op het niveau van bewegingskinematica. Deze conceptie omvat echter verschillende problemen: Ten eerste, het beperkt de studie van wil tot eerder eenvoudige bewegingen zoals het indrukken van toetsen. Ten tweede, het is redelijk controversieel in welke mate interne factoren experimenteel kunnen achterhaald worden (Nachev & Husain, 2010). Ten slotte bestaat er in stijgende mate consensus dat, in de dagdagelijkse ervaring, acties nooit exclusief vrijwillig of door externe stimuli gedreven zijn, maar eerder op een continuüm van “reflexiviteit” variëren (zie Krieghoff, Waszak, Prinz, & Brass, 2011). Dus, het verschil tussen endogene

en exogene oorzaken kan vervagen, vooral wanneer gelijkaardige acties herhaaldelijk worden uitgevoerd, zoals in de meeste psychologische experimenten.

Een andere uitdaging in de studie van wil is het design van geschikte experimentele paradigma's. Wilsmatig gedrag wordt gekenmerkt door het vertrouwen op endogene, in plaats van exogene determinanten (Haggard, 2008). Dus, elke instructie van de experimentator die specifieert hoe de participant zich zou moet gedragen trekt onvermijdelijk in vraag in welke mate daaropvolgend gedrag nog als vrijwillig kan beschouwd worden. Echter, op een zeker niveau is instructie eenvoudigweg noodzakelijk om betekenisvolle data te verkrijgen die gekwantificeerd en vergeleken kunnen worden tussen participanten. De meest voorkomende strategie in de literatuur is om onvolledige instructies te geven, waarin de invulling van bepaalde bewegingsparameters (nl., de selectie of de timing van een beweging) overgelaten wordt aan de participant (Haggard & Eimer, 1999; Krieghoff, Brass, Prinz, & Waszak, 2009; Libet, Gleason, Wright, & Pearl, 1983). Hoewel zulke "gaten" in de taakinstructie zeker beroep doen op interne generatie, ze beperken wil tot zeer specifieke aspecten van het beslissingsproces.

Een laatste belangrijk probleem in de studie van wil betreft motivationele aspecten. Wilsmatig gedrag is typisch gericht op een gewenste uitkomst (zie Haggard, 2008). In andere woorden, we ageren of onderdrukken een actie omdat we een bepaalde (korte termijn of lange termijn) verandering in onze omgeving teweeg willen brengen. Dus, paradigma's gericht op het informeren van wil zouden experimentele contexten moeten creëren waarin zulk gedrag niet enkel endogeen is gegenereerd, maar ook gemotiveerd (in plaats van bv. kiezen tussen het indrukken van linker –of rechtertoetsen). Samengevat illustreren deze overwegingen de obstakels in de studie van wil als een wetenschappelijk construct; men wordt geconfronteerd met een onvermijdbare afweging tussen interne en ecologische validiteit, een kwestie waarop verschillende keren wordt teruggekomen in deze thesis.

Wil bestuderen

Ondanks deze uitdagingen zijn verschillende effectieve strategieën gebruikt om de hierboven genoemde problemen te omzeilen en wil te meten op een wetenschappelijk aanvaardbare en ecologisch valide manier. In wat volgt zal ik kort drie

strategieën beschrijven die in het bijzonder informatief zijn gebleken. De eerste is om indirecte of impliciete metingen van wilsprocessen te meten in plaats van te vertrouwen op introspectie. Bijvoorbeeld, de perceptuele aantrekking tussen een vrijwillige actie en het sensorisch resultaat daarvan is gevalideerd als een methode om het gevoel van agency te meten (nl., de ervaring van het gevolg van een actie te veroorzaken, zie Moore & Obhi, 2012 en latere secties). Impliciete metingen zijn op die manier direct observeerbare en kwantificeerbare variabelen die op een manier in relatie staan met het onobserveerbare proces van wil waarin we geïnteresseerd zijn. Een tweede strategie is om functionele hersenbeeldvorming toe te passen terwijl participanten vrijwillig gedrag uitvoeren. Neurobeeldvorming maakt metingen van observeerbare correlaten van wilsprocessen in de menselijke hersenen mogelijk, en is daardoor een andere manier om bepaalde vrijwillige daden te materialiseren en te kwantificeren. Onderzoek langs deze lijnen heeft aangetoond dat endogene controle van gedrag steunt op de mediale frontale cortex, terwijl de contextuele sturing van gedrag op basis van externe informatie sterker berust op het laterale deel van de frontaalkwab (Amodio & Frith, 2006; Goldberg, 1985; Passingham, Bengtsson, & Lau, 2010; Rushworth, Walton, Kennerley, & Bannerman, 2004). Bovendien zijn recente neurobeeldvormingsstudies begonnen aan het ontbinden van het brede concept van wil in verschillende deelcomponenten op basis van hun neurale handtekening (Brass & Haggard, 2008; Brass, Lynn, Demanet, & Rigoni, 2013; Haggard, 2008; Krieghoff et al., 2009). Echter, ondanks de grote aantrekking van deze methode moet één beperking in gedachten gehouden worden: deze methode steunt meestal cruciaal op omgekeerde inferentie, nl. de interferentie van een psychologisch proces gebaseerd op de activiteit van een bepaald hersengebied, hetgeen deductief invalide en mogelijks misleidend is (Poldrack, 2006). Een laatste strategie is om neuropsychiatrische patiënten te bestuderen van wie bepaalde componenten van wilscontrole fout lopen (vb., Brandt, Lynn, Obst, Brass, & Münchau, 2014). Bijvoorbeeld, anarchistisch-handsyndroom, dat kan resulteren van hersenletsels in het pre-supplementaire motorgebied, is gekenmerkt door de onmogelijkheid om een actietendens te onderdrukken eens die uitgelokt is door een externe stimulus, ondanks de subjectieve ervaring dit te doen (bv., Della Sala, Marchetti, & Spinnler, 1991; Kritikos, Breen, & Mattingley, 2005; Pacherie, 2007). Zulke bevindingen zijn zeer informatief over de fenomenologie, de neurale basis, en de

functionele mechanismen die aanleiding geven tot vrijwillig gedrag en ervaringen. Belangrijk is dat elke van deze strategieën verschillende sterktes en beperkingen hebben, en een tevredenstellende wetenschap van wil vereist een integratie van deze methoden.

Andere onderzoekslijnen zijn minder gefocust op de architectuur van wil, maar eerder op de identificatie van factoren die de mogelijkheid of de motivatie om wilscontrole uit te voeren. Bijvoorbeeld, een reeks studies heeft het idee onderzocht dat wil berust op een beperkte bron van zelfregulatie, en dat de uitputting daarvan (door inspannende acties van zelfcontrole) resulteert in een storing van wilscontrole op daaropvolgende taken (*ego-depletie*, zie Baumeister, Bratslavsky, Muraven, & Tice, 1998). Andere studies hebben gefocust op de invloed van complexe overtuigingen op gedragscontrole. Meer bepaald, overtuigingen over het concept "vrije wil" hebben meer en meer aandacht gekregen; het is aangetoond dat het ondermijnen van deze overtuigingen een substantiële impact heeft op gedrag in sociale contexten (bv., Baumeister, Masicampo, & Dewall, 2009; Vohs & Schooler, 2008) en zelfs op basismotorcontroleprocessen (Rigoni, Kühn, Sartori, & Brass, 2011), hoewel de onderliggende mechanismen nog niet begrepen zijn.

Samenvattend werden er verschillende manieren ontwikkeld om wil te bestuderen, die grotendeels kunnen onderverdeeld worden in (i) onderzoek dat poogt bepaalde componenten van wil te onderzoeken, en (ii) onderzoek dat tracht de factoren te identificeren die de uitoefening van wilscontrole beïnvloeden. Dit onderscheid is ook toepasbaar op de structuur van deze thesis. De eerste onderzoekslijn behandelt de neurale basis van inhibitoire zelfregulatie, met als doel een ecologisch valide manier te bieden om deze capaciteit te onderzoeken. Daarna, in het tweede deel van mijn thesis, is mijn doel om factoren te identificeren die de mogelijkheid om wilscontrole uit te oefenen beïnvloeden.

Deel I: Wil als impulscontrole

Een manier waarop vrijwillig gedrag kan worden gedefinieerd is door de mogelijkheid het te onderdrukken (Passingham, 1993). We ervaren elke dag een brede variëteit van impulsen, sommige zijn gedreven door stimuli, sommige zijn intern gegenereerd. Echter, deze impulsen zijn niet altijd overeenstemmend met lange termijn

doelen of maatschappelijke normen. Bijgevolg is de mogelijkheid om gedragsimpulsen te controleren en te negeren kenmerkend voor adaptief gedrag in complexe sociale systemen.

Om deze mogelijkheid verder op een ecologisch valide manier te onderzoeken stel ik in **Hoofdstuk 1** een nieuw paradigma voor dat gedragsinhibitie in de context van pijnvermijdingsgedrag onderzoekt. Participanten kregen thermale pijnstimulatie aan de binnenzijde van de polsen van afwisselend het rechter –en linker hand en konden de stimulatie beëindigen door een toets in te drukken met het niet-gestimuleerde hand. Op sommige trials werd de beslissing om te reageren of deze respons te onderdrukken aangegeven door een externe cue, terwijl op andere trials participanten zelf vrij konden kiezen tussen beide opties. Het voordeel van dit paradigma is dat het een sterke drang induceert om te reageren in elke participant, in elke trial. Responsinhibitie in dit paradigma spiegelt dus meer realistische gevallen van zelfcontrole dan in vroegere paradigma's. Opvallend is dat in deze motivationeel opvallende condities inhibitie berust op een gemeenschappelijke activatie van de hersennetwerken die betrokken zijn in zowel gedrag dat door externe factoren wordt geleid én intentionele inhibitie. Deze bevinding trekt de oorspronkelijke aanname dat deze netwerken onafhankelijke functies dienen in twijfel en suggereert dat deze mogelijks samen opereren in de context van echte zelfcontrole. In **Hoofdstuk 2** integreer ik deze en andere neurobeeldvormingsbevindingen in een nieuwe theoretisch uiteenzetting van de rol van de dorsale frontomediane cortex in intentionele inhibitie. Op basis van een aantal studies die deze regio gelinkt hebben aan non-motor en zelf-referentiële processen argumenteer ik dat de bijdrage van deze regio aan zelfregulatie beter beschouwd kan worden als een complexe onthechting van intenties en driften, en niet zozeer als een vetosignaal dat motorplannen annuleert in de strikte zin.

Deel II: Determinanten van vrijwillig gedrag

Het tweede deel van mijn thesis verplaatst de focus van het meten en kenmerken van wil zelf naar het omschrijven van factoren die de mogelijkheid en/of de motivatie om wilscontrole uit te oefenen beïnvloeden. Volgend op de bepaling van de mechanismen betrokken in de inhibitie van pijnvermijding gebruiken **Hoofdstukken 3 en 4** een aangepaste versie van hetzelfde paradigma. **Hoofdstuk 3** onderzoekt de

relatie tussen inhibitie-gerelateerde hersenactiviteit en individuele verschillen in mogelijkheden tot zelfregulatie (zowel op het niveau van toestand als trek). Resultaten tonen dat een hoge mate van zelfregulatie als trek geassocieerd is met sterkere activatie van het inhibitienetwerk, zoals gedefinieerd in het vorige hoofdstuk, terwijl geen effecten van manipulaties van de toestand werden gevonden. **Hoofdstuk 4** onderzoekt overtuigingen over wil en hun impact op gedragscontrole door te onderzoeken of het verzwakken van overtuigingen over vrije wil de intentionele betrokkenheid in zelfcontrole beïnvloedt. Een aantal studies geeft aan dat dit het geval zou kunnen zijn. Echter, de effecten worden enkel gevonden in een deel van de sample, nl. in participanten die voldoende pijn ervoeren om zelfcontrole nodig te hebben. **Hoofdstuk 5** is gericht op het onthullen hoe overtuigingen over vrije wil verschillende gedragsparameters kunnen beïnvloeden, van het gereedheidspotentiaal tot complexe sociale gedragingen. Meer bepaald, het test de hypothese dat het verzwakken van overtuigingen over vrije wil het gevoel van agency vermindert, wat grotendeels bevestigd werd door de data.

Referenties

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–277.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252–1265.

Baumeister, R. F., Masicampo, E. J., & Dewall, C. N. (2009). Prosocial benefits of feeling free: disbelief in free will increases aggression and reduces helpfulness. *Personality & Social Psychology Bulletin*, *35*(2), 260–8.

Brandt, V. C., Lynn, M. T., Obst, M., Brass, M., & Münchau, a. (2014). Visual feedback of own tics increases tic frequency in patients with Tourette's syndrome. *Cognitive Neuroscience*, (October 2014), 1–7.

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *14*, 319–325.

Brass, M., Lynn, M. T., Demanet, J., & Rigoni, D. (2013). Imaging volition: what the brain can tell us about the will. *Experimental Brain Research, 229*(3), 301-312.

Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, *29*, 1113–1117.

Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behavioral and Brain Sciences*, *8*, 567–588.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews. Neuroscience*, *9*(12), 934–46.

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

Harleß, E. (1861). Der Apparat des Willens. *Zeitschrift Für Philosophie Und Philosophische Kritik*, *38*, 499 – 507.

Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *Quarterly Journal of Experimental Psychology*, *60*(11), 1540–54.

Herwig, A., & Waszak, F. (2009). Intention and attention in ideomotor learning. *Quarterly Journal of Experimental Psychology*, *62*(2), 219–27.

James, W. (1890). *The principles of psychology (Vols. 1 & 2)*. New York: Dover.

Krieghoff, V., Brass, M., Prinz, W., & Waszak, F. (2009). Dissociating what and when of intentional actions. *Frontiers in Human Neuroscience*, *3*(February), 3.

Krieghoff, V., Waszak, F., Prinz, W., & Brass, M. (2011). Neural and behavioral correlates of intentional actions. *Neuropsychologia*, *49*(5), 767–76.

Kritikos, A., Breen, N., & Mattingley, J. B. (2005). Anarchic hand syndrome: Bimanual coordination and sensitivity to irrelevant information in unimanual reaches. *Cognitive Brain Research*, *24*, 634–647.

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*, *106 (3)*, 623–642.

Lotze, R. H. (1852). *Medizinische Psychologie oder Physiologie der Seele.* Leipzig, Germany: Wedmann'sche Buchhandlung.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Consciousness and Cognition*, *21*(1), 546–61.

Morsella, E., Molapour, T., & Lynn, M. T. (2013). The three pillars of volition: phenomenal states, ideomotor processing, and the skeletal muscle system. In H. S. Terrace & J. Metcalfe (Eds.), *Joint Action and Agency*. New York: Oxford University Press.

Nachev, P., & Husain, M. (2010). Action and the fallacy of the "internal": Comment on Passingham et al. *Trends in Cognitive Sciences*, *14*(5), 192 – 193.

Pacherie, E. (2007). The anarchic hand syndrome and utilization behavior: A window onto agentive self-awareness. *Functional Neurology*, *22*, 211–217.

Passingham, R. E. (1993). *The frontal lobes and voluntary action. Oxford psychology series, No. 21*. New York, NY: Oxford University Press.

Passingham, R. E., Bengtsson, S. L., & Lau, H. C. (2010). Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14*, 16–21.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.

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*, *22*(5), 613–8.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences, 8*(9), 410-417.

Shadlen, M. N., & Gold, J. L. (2004). The neurophysiology of decision making as a window on cognition. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III* (pp. 1229 – 1243). MIT Press.

Vohs, K. D., & Schooler, J. W. (2008). The Value of Believing in Free. *Psychological Science*, *19*(1), 49–54.

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), 346–56.