Metacognition in attention control and skilled action
—mindfulness and flow
PhD Thesis
Victor Lange

Metacognition in attention control and skilled action
—mindfulness and flow

Supervisor: Thor Grünbaum
Submitted on: 14.12.2023
Name of department: Section for Philosophy, Department of Communication, University of Copenhagen

Author: Victor Lange

Title: Metacognition in attention control and skilled action—mindfulness and flow

Supervisor: Thor Grünbaum

Submitted on: 14.12.2023

Word count: 63,338 (including references)

Funding: The present PhD project was a part of the research project “The Functional Role of Perception of Movement,” funded by a grant from the Independent Danish Research Council, DFF FKK, grant number 0132-00141B.
Table of contents

Acknowledgements ............................................................................................................. 5
Coauthors .......................................................................................................................... 7
Abstracts in English and Danish ....................................................................................... 10
Article overview ............................................................................................................... 12
Introduction ........................................................................................................................ 14
Article 1: Decentering and attention ................................................................................. 53
Article 2: Transparency and the Mindfulness opacity hypothesis ...................................... 81
Article 3: A reductive account of mindfulness as metacognitive control ............................. 104
Article 4: Flow and experts ............................................................................................... 129
Concluding remarks and further research ......................................................................... 152
Acknowledgments

I would like to thank Thor Grünbaum, who has been a truly dedicated and incredible supervisor. Thor has shown a deep interest, respect, and faith in my work. Moreover, he has had great ambitions on behalf of the present thesis, meaning that he has been the most critical reader of the articles. This mixture of interest, curiosity, ambition, and critical assessment has been profoundly fruitful to my work. Aside from that, Thor has been very helpful in multiple other respects, such as in arranging research stays and participation in workshops and conferences. He has been very open-minded and supportive in letting me pursue the activities that interested me. I have really enjoyed our collaboration and learned a lot from it. I am grateful for the many great and fun conversations we have had over a beer or dinner over the years.

I would also like to thank Mark Schram Christensen, who directed the research project “The Functional Role of Perception of Movement,” of which the present PhD thesis is a part. I have learned a great deal through working with Mark—for example, through the coauthoring of a neuroscientific paper on flow with Mark and Frederik Merrild Noorpil. Mark has supported me in pursuing my research interest, and I have really enjoyed our conversations on topics such as flow and skill. Relatedly, I would like to thank Ida Marie Brandt, who has written her PhD project on the same research project. Discussing topics such as action, attention, consciousness, and control with Ida Marie and Mark has been very interesting to me. These discussions have offered me a window to understand how motor neuroscience investigates such phenomena.

Next, I would like to thank Elisabeth Pacherie and Denis Buehler for welcoming me to Institut Jean Nicod at Ecole Normale Supérieure Paris. I really enjoyed my stay there, and I learned a lot. I find the work of Elisabeth and Denis deeply inspiring and interesting. I also extend my thanks to Ophelia Deroy for welcoming me into the Cognition, Value, and Behavior group at Ludwig-Maximilians-Universität in Munich. I learned a great deal from that stay—Ophelia’s views on philosophy, science communication, and public engagement were really inspiring. Also, I would like to express my gratitude to Rick Repetti for offering very interesting comments on some of the articles in this PhD and for taking the role of external evaluator in my mid-way workshop. I really appreciate it. On a similar note, my thanks go to Søren Overgaard for being the other evaluator in my mid-way workshop. Moreover, I would like to thank the members of the CoInAct group at the University of Copenhagen, especially Astrid Kjær Reng, Gustav Strægaard Vinding Runge, Markus Grenaa Giessing, and Frederik Tollerup Junker, for many inspiring conversations on various issues in philosophy of mind.
Next, I would like to acknowledge people outside of my official work environment. First, I would like to thank my wife Sidsel Marie for being the most wonderful and amazing person in the world. Throughout this three-year PhD, Sidsel Marie and I have lived in seven different apartments in five different cities. Wherever we have stayed, I have ended up with a sense of home and belonging (you know, home is where the heart is, as Elvis sings). Also many thanks for countless conversations on the science of contemplative practice, psychological illness, and the possibility of spirituality without religion. Those conversations have been profoundly inspiring, and they have resulted in publishing papers and doing research together. I find Sidsel Marie’s anthropological and existential perspectives deeply exciting and have learned so much from these approaches to the human mind.

Moreover, I would like to extend my thanks to several other philosophers and friends with whom I have had conversations throughout the years, even though we have not been a part of the same official work environment. I would like to thank Andreas von Achen for numerous discussions on the relationship between philosophy and natural sciences and many conversations on the relevance of philosophy in political decision-making. I am also grateful to Vincent Ringgaard for many fantastic conversations on Buddhism, Stoicism, meditation, martial arts, and philosophy as a way of life. Next, I would like to thank Susanne Henningsson for many conversations on the relationship between philosophy and neuroscience and for great discussions on the importance of philosophy and artistic activities to human well-being. Likewise, my thanks go to Helene Borregaard Stephensen for many inspiring conversations on the importance of philosophical research and practice in relation psychiatry. These conversations have touched upon the benefits and limits of approaches in both the analytic and the more “continental” traditions of philosophy. I have learned a great deal from these exchanges of ideas.

Next, I would like to express my gratitude to my dear parents for their interest in my work and for being so incredibly helpful. Thanks also go to my sister, Astrid, for her deep interest in all my projects. I would like to thank Knaster for being so interested in my work and for housing me for small writing retreats. Thanks also to my many wonderful friends for their curiosity in my work: K, Eva, Mill, Tina, Hinni, Jonas, Rebecca, Julius, Mau, Tuborgh, C, Niels, and many, many more. I am also grateful to Marianne and Peter for always being so genuinely interested in my academic projects and life. Many thanks to Lene, Per, Anne Stine, and Andreas for being so extremely helpful. And finally, thank you to Lea and Bernard for reminding me of how philosophy can be a passion that burns through hard times.
Coauthors

Thor Grünbaum coauthors Article 2. See the following page for the official documentation.
### Co-author statement

**PhD student**  
Victor Lange

**Date of birth**  
15.04.1993

**Faculty (Department)**  
Section for Philosophy, Department of Communication

"Attribution of authorship should in general be based on criteria a-d adopted from the Vancouver guidelines, and all individuals who meet these criteria should be recognized as authors:

A. Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, and
B. drafting the work or revising it critically for important intellectual content, and
C. final approval of the version to be published, and
D. agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."

**Article/paper/chapter/manuscript**

This co-authorship declaration applies to the following:

*Title*  
Transparency and the Mindfulness opacity hypothesis

*Author(s)*  
Victor Lange & Thor Grünbaum

**Journal**  
Philosophical Quarterly

**Contributions to the paper/manuscript made by the PhD student**

**What was the role of the PhD student in designing the study?**

The PhD student was the prime mover in conceptualising and designing the study.

**How did the PhD student participate in data collection and/or development of theory?**

The PhD student was the prime mover in developing the theory, surveying the literature, designing the structure, and developing the argument.

**Which part of the manuscript did the PhD student write or contribute to?**
The PhD student wrote all sections of the first draft, as well as subsequent drafts based on feedback from the co-author.

Did the PhD student read and comment on the final manuscript?

Yes.

Signatures

If an article/paper/chapter/manuscript is written in collaboration with three or less researchers (including the PhD student), all researchers must sign the statement. However, if an article has more than three authors the statement may be signed by a representative sample, cf. article 12, section 4 and 5 of the Ministerial Order No. 1039, 27 August 2013. A representative sample consists of minimum three authors, which is comprised of the first author, the corresponding author, the senior author, and 1-2 authors (preferably international/non-supervisor authors).

By their signature, the authors agree that the article/paper/chapter/manuscript will be included as a part of the PhD thesis made by the PhD student mentioned above.

Date 10/12/2023 Name Thor Grünbaum Signature

Date Name Signature

Date Name Signature
Abstract

This PhD thesis examines the role of metacognition in attention control and skilled action. Metacognition refers to those mental processes of an agent that monitor, evaluate, or regulate her other mental processes. The articles explore how metacognitive processes support attention control and skill execution with a focus on two specific psychological phenomena, namely mindfulness and flow. The introduction of the thesis provides an overview of the relevant philosophical literature, clarifying how Articles 1–4 contribute to these research fields. The introduction focuses on three topics, namely metacognition, attention, and action. It also provides some considerations on how philosophy can contribute to interdisciplinary research, notably clinical research and practice.

Article 1 concerns the phenomenon of decentering. Decentering concerns an agent’s ability to move out of immersion in her own mental state(s). It is an ability that is at the center of contemporary and mindfulness-based cognitive psychotherapies, known as third-wave cognitive behavior therapies. The article provides a novel account of decentering as a complex mental operation composed of simultaneous introspection and detachment. It argues that the popular selection for action theory of attention cannot account for this nature of decentering.

Article 2 discusses the so-called Transparency thesis, well known in philosophy of consciousness and perception. In recent years, multiple mindfulness researchers have claimed that mindfulness introspection is in conflict with this thesis. Article 2 labels this claim the Mindfulness opacity hypothesis. Article 2 develops the hypothesis in various philosophical respects, arguing that it is well motivated both philosophically and empirically.

Article 3 targets a central problem in the science of mindfulness. This is the problem of whether mindfulness involves an essential and categorically demarcated psychological capacity. The most promising proposal is that this capacity is a form of metacognitive control. Yet, Article 3 argues that previous accounts of such metacognitive control all fail. Hence, the article defends a novel account, arguing that this provides new insights into mindfulness.

Article 4 discusses the phenomenon of flow, i.e., states where agents execute an activity with fluency, deep focus, and some feelings of enjoyment or meaning. The philosophical position known as the Automatic view of expertise claims that flow states are characteristic to athletic expertise and that this shows why experts do not consciously supervise their activity when performing optimally. Article 4 argues that this view is deeply mistaken in the light of the current psychology of flow, concerning attention and metacognition.
Résumé

Denne Ph.d.-afhandling undersøger rollen af metakognition i opmærksomhedskontrol og udførelsen af færdigheder. Metakognition omhandler de mentale processer hos en agent der montererer, evaluerer eller regulerer andre mentale processer hos selvsamme agent. Artiklerne i denne afhandling udforsker, hvorledes sådanne metakognitive processer er involveret i to specifikke psykologiske fænomener, nemlig mindfulness og flow. Afhandlingens introduktion giver et overblik over den relevante filosofiske litteratur, samt en række betragtninger over hvorledes filosofi kan bidrage til interdisiplinære forskningsområder—i særlighed forskning og praksis i klinisk psykologi.


**Article overview**

**Article 1: Decentering and attention**
Clinical psychologists describe decentering as the mental operation in which a subject “moves out” of immersion in a mental state. Such decentering is philosophically puzzling. It involves a subject attending to her mental state to distance herself from it. That is, she attends to the state to make it less determining of her processing. This paper provides a philosophical explanation of the nature of decentering. It analyzes decentering as a complex mental operation composed of two suboperations: introspection and detachment. Drawing on this analysis, the paper argues that decentering involves certain dynamics of attention and attention control that pose an important challenge to a selection for action theory of attention. This challenge concerns adequately describing the workings of detachment in decentering. The paper discusses how a selection for action theory might reply, yet it argues that all the available replies involve unattractive aspects. The paper closes with broader perspectives, suggesting that decentering might also pose a puzzle for other theories of attention.


**Article 2: Transparency and the Mindfulness opacity hypothesis**
Many philosophers endorse the Transparency thesis, the claim that by introspection one cannot become aware of one’s experience. Recently, some authors have suggested that the transparency thesis is challenged by introspective states reached under mindfulness. We label this the Mindfulness opacity hypothesis. The present paper develops the hypothesis in important new ways. First, we motivate the hypothesis by drawing on recent clinical psychology and the cognitive science of mindfulness. Second, we develop it by describing the implied shift in experiential perspective, the scope of introspectable qualities, and the level of skill. Third, we defend the hypothesis against various philosophical arguments. We conclude that the Mindfulness opacity hypothesis is empirically and theoretically well motivated and supported.

Thor Grünbaum coauthored Article 2.

**Article 3: A reductive account of mindfulness as metacognitive control**

Mindfulness is a large research field, involving disciplines such as philosophy, cognitive psychology, psychiatry, neuroscience, and Buddhist studies. Despite this widespread interest, one question remains unanswered: Is there a psychological capacity that is essential to mindfulness and which demarcates mindfulness from most other mental activities? The most promising idea is that mindfulness is a special form of metacognitive control. Yet, I argue that current proposals on how to conceptualize such metacognitive control fail. Instead, I propose a novel account of the metacognitive control of mindfulness, drawing on the idea of so-called metacognitive goals. This account allows us to make sense of the explicit self-awareness and self-regulation involved in mindfulness and to separate mindfulness from exercises of more ordinary cognitive control. According to this account, metacognitive control is only a necessary and not a sufficient condition for mindfulness. Finally, I argue that the account motivates two theses on the nature of mindfulness, namely that we can reduce the metacognitive control of mindfulness to other psychological capacities and that this control is a form of mental action.

**Article 4: Flow and experts**

Philosophers frequently refer to flow when discussing skill and expertise. This is especially clear when considering the “flow argument” advocated by the Automatic view. Although the Automatic view is a minority position today, many philosophers and scientists endorse the flow argument or central parts of it. The present paper argues that the argument is implausible in the light of current psychological research. It proceeds in two steps. First, the paper argues that psychological research does not suggest that flow is the only type of psychological state characteristic of optimal expert performance. Experts undergo different states, notably the state of clutch. Second, the paper argues that psychological research does not suggest that flow states are devoid of conscious supervision in the form of conscious control, self-awareness, or task-related thoughts. Experts execute multiple conscious operations involved in attention control, motivational self-talk, and certain psychological mindsets while being in flow. These considerations oppose the two premises of the flow argument. Moreover, they correct misunderstandings of flow in the broader philosophical and scientific literature.
1. Introduction

For much of our waking time, our mental life concerns our external environment. We see the bike lanes full of people in front of us; we hear the sound of sirens; we smell the freshly brewed coffee; we tactiley feel the texture of the table; we reflect upon where to go for dinner; we try to remember the meeting schedule for next week; we think about calling a friend; we mind-wander about where to travel this summer. Yet, sometimes, our mental life does not concern the world around us in this way. Instead, it concerns itself. For example, we suddenly become aware that our mind has wandered away from the paper we were supposed to read; we doubt that our memory of a past event is accurate; we have the feeling that our answer on a test is mistaken; we reflect upon why we have certain emotions. Our mental life turns toward itself in such cases. We have mental processes (feelings, judgements, ideas, awareness, etc.) about our own mental processes. Philosophers, psychologists, and cognitive scientists commonly refer to such instances as occurrences of metacognition. Even though the exact definition of metacognition is philosophically controversial (Proust, 2013), many authors accept the rough characterization that metacognition concerns those mental processes of an agent that monitor, evaluate, or regulate other mental processes of the same agent.

The concept of metacognition has long been widely used in the empirical sciences. In recent years, an increasing number of philosophers have adopted the concept into discussions in philosophy of mind and philosophical psychology. The four articles in this PhD thesis are a contribution to these discussions. The topic of the thesis is the role of metacognition in attention control and skilled action. This is a broad subject matter. To narrow its focus, the thesis examines two specific phenomena that involve this broader topic in interesting ways. These are the phenomena of mindfulness and flow. Both are subject to widespread and interdisciplinary research (Baminiwatta & Solangaarachchi, 2021; Abuhamdeh, 2020). They are also of clinical interest, especially mindfulness. Mindfulness is a core component in contemporary cognitive-behavior psychotherapies (Rosenkranz et al., 2019). Lastly, there is much public interest in mindfulness and flow. There are thousands of popular writings on these phenomena. Mindfulness and flow have been some of the most central ideas in the last 20 years of self-help literature (Kabat-Zinn, 2005; Csikszentmihalhi, 2020).

Article 1 of this thesis provides a philosophical analysis of so-called decentering. Decentering concerns an agent’s ability to move out of immersion in her own mental state(s). This ability is central to the contemporary and mindfulness-based cognitive psychotherapies,
known as third-wave cognitive-behavior psychotherapies. Studies suggest these therapies to be among the most efficient psychotherapeutic protocols for treating a wide range of psychological illnesses (Dimidjian et al., 2016). Article 1 aims to contribute to the philosophical literature in two ways. First, it provides the first dedicated, in-depth philosophical analysis of decentering. The paper argues that decentering is a complex mental operation in which agents execute two suboperations simultaneously, namely, introspection and detachment. In particular, the paper provides an extensive analysis of detachment. It argues that such detachment is a matter of agents establishing certain kinds of stances towards their own mental states. These are the stances of disidentification and nonreactivity. Second, drawing on this analysis, the paper argues that decentering is a problematic case for the prominent philosophical and psychological theory of attention known as the selection for action theory.

Article 2 concerns a well-known idea in philosophy of mind known as the Transparency thesis. This thesis claims that when we introspect our experiences, we can only become aware of the representational objects involved in our experiences and the properties of these objects. We cannot become aware of any properties that do not appear as such representational objects or properties of them. The transparency thesis has been important in philosophy of mind, both as a thesis about perception and introspective awareness and as a cornerstone in metaphysical theories of consciousness, notably Representationalism (Lycan, 2023; Tye, 2002). Article 2 presents a novel philosophical criticism of the transparency thesis by considering a widespread idea among mindfulness researchers. This is the idea that mindfulness introspection enables states of introspective awareness that are in conflict with the transparency thesis. The article labels this idea the Mindfulness opacity hypothesis. The article develops the hypothesis, drawing on relevant philosophical work and empirical studies on mindfulness. It argues that the hypothesis is well motivated both philosophically and empirically.

Article 3 focuses upon a specific problem in the science of mindfulness. Mindfulness researchers commonly discuss this question: Is there a psychological capacity that is essential to mindfulness and which demarcates mindfulness from most other mental activities? This is an important research question. If we can identify such a capacity, we might be able to better explain various findings on mindfulness, such as the relations between mindfulness practice and increases in attention control, cognitive functioning, emotion regulation, and so forth. Further, identifying this capacity might also enable us to enhance mindfulness protocols and provide better instructions to clients. In contemporary mindfulness research, the dominant proposal is that mindfulness involves a form of metacognition, more precisely, a form of
metacognitive control (Dorjee, 2016; Jankowski & Holas, 2014). Article 3 argues that although this basic idea is plausible, current proposals on how to define such metacognitive control all fail. Instead, Article 3 offers a novel account of the metacognitive control of mindfulness. This account draws on the idea of metacognitive goals. The article provides several considerations in favor of the new account and uses it to motivate two novel theses about mindfulness and its relation to reductionism and mental action.

Taken together, Articles 1, 2, and 3 offer a philosophical framework for understanding mindfulness practice. They argue that metacognitive dynamics are integral to this practice. These dynamics are involved in attention control and other regulative dynamics. In sum, the framework proposes that we consider mindfulness as a skilled mental action.

Article 4 examines the relation between flow and athletic expertise. This connects to broader issues in the philosophy of skill and expertise. For many years, the so-called Automatic view has been dominant among philosophers and psychologists (Dreyfus, 2007a, 2007b). This position claims that when athletic experts perform optimally, it is a matter of automatized and embodied processes running freely without any conscious supervision. That is, experts do not consciously control their activities, are not self-aware of these activities, and do not think while acting. Flow has been a key element in motivating the Automatic view. Article 4 makes two key contributions to this discussion. First, it makes explicit an often implicitly assumed argument for automaticity in terms of flow. The article labels this the flow argument. Second, the article carefully examines each premise in the flow argument. It argues that they are all implausible in the light of current psychological research on flow. Given that the flow argument has been widely assumed and accepted, including by authors that otherwise do not endorse the Automatic view, this critical examination also corrects misunderstandings of flow in the broader philosophical and psychological literature.

During the writing of this PhD thesis, I have also coauthored a neuroscientific paper on flow entitled “Flow experience reflected in increased posterior and frontal beta oscillatory power and lateral gamma power measured in a physical and a computer game.” The two other authors of this paper are Frederik Merrild Noorpil and Mark Schram Christensen. The paper investigates EEG signatures for flow experiences in two tasks (i.e., a physical task and a computer game). Frederik and Mark designed and ran the study, did the various analyses of the data material, and wrote all sections concerning this data and the analyses of it. My contribution was to analyze and discuss the results in relation to the current psychological literature on flow, which involved examining the implications for contemporary theories and conceptualizations of flow. This paper is currently under review. I have not included it in this PhD thesis due to
the length of the thesis and the fact that the paper does not directly target philosophical debates on flow, skill, or expertise.

Articles 1–4 aim to provide novel arguments and analyses to discussions in philosophical psychology. The articles concern a wide range of specialized discussions in philosophy of metacognition, philosophy of attention, philosophy of action, and philosophy of meditation. This introduction contextualizes these contributions by showing how they are related to ongoing philosophical discussions. The introduction is organized around the three key concepts of the present PhD thesis, namely, metacognition, attention, and action. Lastly, the final section of the introduction offers methodological considerations about how philosophical work, like that of Articles 1–4, might contribute to interdisciplinary endeavors. This also concerns contributions to clinical practice.¹

1.1 Personal-level and sub-personal-level processes
Initially, let me clarify the distinction between personal-level and sub-personal-level mental processes (see Drayson, 2014, for a good introduction). This distinction appears continuously throughout the following sections. Originally, Dennett (1969) introduced the distinction to stress a difference between two types of psychological explanations. Roughly, Dennett (1969) states that personal-level explanations refer to the level of people (their activities, perceptions, conscious thoughts, etc.), while sub-personal-level explanations refer to events or mechanisms in the brain and nervous system of people (levels of certain neurotransmitters, electrical signals, computational systems, etc.). It is not always clear how Dennett views the relationship between these two types of explanations. In some places, Dennett (1987) seems to view personal-level explanations as merely instrumental or prescientific explanations of cognitive systems, which sub-personal explanations should replace when possible. In other places, one might interpret him as expressing a less reductionist position (1991). Nevertheless, since Dennett’s introduction, many other philosophers have come to adopt the distinction, often with the purpose of defending or developing their own positions with respect to explaining the mind (e.g., Fodor, 1975; Lycan, 1987). However, as Drayson (2014) argues, these adaptations were actually not neutral. The authors added several further aspects to the distinction, making it

¹ Articles 1–3 continue a line of research that I began in my master’s dissertation (entitled A philosophical analysis of metacognitive attention control). This master’s dissertation also analyzed mindfulness as a form of metacognition, relating it to philosophical discussions of attention and the Transparency thesis. Nonetheless, the PhD thesis makes no use of concrete analyses or text units from the master’s dissertation. The relationship is only on a very abstract level of ideas and not on a substantial level of argumentation or research.
importantly different from the original and seemingly more minimal formulation in Dennett’s work (1969).

The present PhD thesis is not an attempt to interpret Dennett’s original proposal nor to track the history of the distinction in the philosophical literature. For my argument, the exact way of spelling out the distinction is not crucial. Nonetheless, I have decided to use the distinction as an organizing principle in these following introductory sections and their reviews of the philosophical literature on metacognition, attention, and action. The distinction, despite its confusing and divergent uses in the philosophical literature, allows us to capture important differences and aspects of the current philosophical debates on these topics. Let me clarify how the following sections approach the distinction. Overall, notice that I use the distinction to point to a difference primarily in terms of mental processes, and not explanations (as Dennett originally proposed). The following sections view personal-level mental processes as processes that we can ascribe to the person as a whole, while sub-personal mental processes cannot be ascribed to the person as a whole but only to a sub-system of the person’s mind. With this basic characterization in mind, many authors consider reasoning (Soteriou, 2013) or emotions (Feldman-Barret, 2017) as personal-level mental processes. Conversely, mental processes like early visual processing (Marr, 1982) or fine-grained motor control (Fourneret & Jeannerod, 1998) are often considered sub-personal mental processes. Nonetheless, this characterization still leaves us with the question: What are the exact requirements for such ascription? I have no well-developed position on this matter. However, let me stress that I use the distinction in a way that implies two important differences between personal-level and sub-personal mental processes.

First, when I use the distinction, I imply the following conditional: If a mental process is conscious or consciously accessible through introspection, then it strongly indicates that we can ascribe the process to the agent as a whole (i.e., it is a personal-level process). On the other hand, sub-personal processes are subconscious or nonconscious, and we do not expect them to be accessible to conscious introspection. Hence, ascribing mental processes to persons as a whole is associated with consciousness in this sense. One might respond that this idea is controversial. For example, there are numerous mental processes and states that are personal-level phenomena, yet they are not conscious and perhaps not even consciously accessible through introspection (e.g., implicit biases or attitudes in social cognition: Brownstein & Saul, 2016). I am open to this idea, and it is totally compatible with the above conditional (since it is not a claim about consciousness being necessary but sufficient to indicate categorization of a mental process as personal-level). Further, note that this is a claim about the mental process
being conscious or open to introspection. As we shall see in §2, many authors think that subpersonal and nonconscious metacognitive processes give rise to conscious signals in the form of affective states. This is different from the present idea, since these authors propose that the relevant metacognitive process unfolds outside of the agent’s consciousness and access of introspection, even though it produces such a conscious signal.

Second, I also imply the following claim when I use the distinction: It is not a necessary requirement for a mental process to have clear and distinct neural implementations for it to qualify as a personal-level mental process. That is, I allow that personal-level mental processes are complex or global in the sense that they involve multiple neural systems (e.g., an emotion of despair might involve interoceptive signals, pre-frontal contextualization, or associations in hippocampus). Further, these processes also exhibit heterogeneity in the sense that they might involve different neural systems in different situations (e.g., despair in one situation might have certain neural signatures that differ from despair in another situation). This is in contrast to subpersonal processes. Most often, we expect that sub-personal processes have clear and distinct neural implementations.2

These two considerations leave many things open. For example, many authors believe that a core characteristic of personal-level processes is that they are governed by norms of rationality, relating to issues of autonomy, control, and reason (Fodor, 1975; Hurley, 1998). I do not wish to discuss these matters here. I only aim to provide a very minimal characterization that spells out the notion involved in the present PhD thesis. It is not an attempt to argue for a general account of the difference between personal-level and sub-personal processes; nor is it an attempt to argue for a certain relationship between these processes. Nevertheless, I think the above approach corresponds to how many authors in philosophical psychology view the distinction, relating to discussions of metacognition (Proust, 2013; Shepherd & Mylopoulos, 2021), attention (Buehler, 2023; Watzl, 2017), and action (Wu, 2016; Brozzo, 2021). This should be sufficient for the use of the distinction in the following literature reviews.

---

2 Notice that I do not endorse the position that only sub-personal processes are proper objects of science (since these have clear neural implementations or computational workings, in contrast to personal-level processes). I am not sympathetic to such a view—and it is clear that it rests on controversial reductionist ideas of scientific practice.
2. Metacognition

Metacognition is a central topic in various research branches, among them cognitive neuroscience (Fleming & Lau, 2014), social neuroscience (Frith, 2012), educational psychology (Hacker et al., 2009), clinical psychology (Wells, 2011), and sports psychology (MacIntyre et al., 2014). Researchers usually classify a mental process of an agent, A, as being metacognition if the mental process monitors, evaluates, or regulates A’s other mental processes. Some authors have been skeptical toward the concept of metacognition, claiming that the concept does not offer anything new to philosophical discussions. They stress that other notions—like self-awareness, self-knowledge, cognitive control, etc.—are clearer and better suited for philosophical debates (Smith, 2014). Yet, in recent years, an increasing number of philosophers have begun introducing and discussing metacognition in relation to a wide range of philosophical topics. Let me review three debates of major importance. This review will show that philosophers have approached metacognition in distinct way. They have considered metacognitive processes to be sub-personal mental processes that produce conscious signals, which agents can use in their personal-level mental operations. The present PhD thesis focuses on another kind of metacognition. It considers types of metacognition that are personal-level mental processes themselves, i.e., conscious psychological processes that we can ascribe to the agent as a whole.

2.1 Epistemic and perception-motor functions

Numerous philosophers have discussed the workings of metacognition with respect to epistemic issues. One line of discussion concerns so-called metacognitive feelings. Building on empirical research on human and animal cognition, Proust (2013) argues that metacognitive feelings are central to processes such as perceptual judgment, decision-making, categorization. Metacognitive feelings involve interaction between two levels of processing in an agent’s mind. Say an agent performs a certain task, such as a perceptual task of determining distance between two objects or a task of matching items in a certain way. This task performance involves the utilization of certain first-order systems. These are the systems that compute a response to the task (notably, the visual systems involved in computing the distance or the memory systems involved in matching items). Proust (2013) argues that the processing of such first-order systems is largely sub-personal to the agent. It is subconscious and distributed to specific subsystems. Yet, despite processing the relevant task sub-personally, the first-order
systems also produce a conscious signal in the form of a feeling. Such feelings are nonconceptual affective signals that indicate epistemic properties of the output of the first-order system. Examples might be a feeling of fluency in processing the task or a feeling of knowing the correct answer without being able to come up with it (as in the tip-of-the-tongue phenomenon: Schwarts, 2006). These feelings are important for the processing of the agent’s meta-level system. The meta-level system does not directly process the relevant task, yet it plays a role in regulating the first-order system’s processing, as in continuing or ending it. The meta-level system uses these feelings as sources for monitoring the success of the first-order system. For example, if the first-order system is accompanied by a feeling of fluency, the meta-level system might consider the output as having a high probability of being correct (although the context is highly decisive for how the meta-level interprets these affective signals of metacognitive feelings: Proust, 2013).

This idea of metacognitive feelings has inspired other lines of discussion. One focus has been on human conscious thinking. Shea (in press) argues that metacognitive feelings play an important role in transitions in conscious thinking. The idea is that transitions in conscious thinking are accompanied by feelings concerning the reliability of the involved steps. These feelings are important to whether we accept our reasoning and its products. We might interpret a feeling of fluency as indicating that the transitions are natural and functioning well. Another, and related, idea is that of concept-metacognition (Shea, 2018, 2020). Concepts are crucial tools for cognition, perception, and action-planning. Shea argues that metacognition about concepts is frequent and important. Such metacognition might take the form of explicit reflection on a concept, yet it might also be a matter of metacognitive feelings. Agents seem to have feelings about the concepts in their repertoire. For example, an agent might have a feeling of confusion about the concept of “electron” when she uses it. Such feelings might indicate that the agent is not fully competent in the application of the concept.

Despite this interest, the idea of metacognitive feelings is controversial. Some authors argue that metacognitive feelings are not metacognitive at all. Carruthers (2011, 2017) argues that the empirical studies that allegedly should support the idea of metacognitive feelings are open to an equally plausible and alternative interpretation. This interpretation suggests that noetic feelings only involve first-order workings. For example, the noetic feelings of fluency or doubt do not stem from a meta-level system receiving a metacognitive feeling. These feelings simply stem from agents that observe or interpret their own situation, for example as in observing their own bodily hesitation in some decision-task and, hereby, generating the feeling of doubt. Another, and more fundamental, criticism objects that the idea of
metacognitive feelings assumes too thin of a notion of metacognition. We can outline this notion as follows: Take a cognitive system, C. If one cognitive sub-system of C, C*, uses a signal to evaluate the success of another cognitive sub-system, C**, then this qualifies as an occurrence of metacognition. Critics argue that this notion is too thin because it implies that metacognition is vastly spread in most operations of cognitive systems, such as the human mind. Take predictive systems like that of the sensorimotor system. Here, the motor control system regulates movement by relying on discrepancies arising from comparing actual sensory inputs with predictive models. This involves one sub-system (the comparator/controller system) evaluating the success of another (the motor production sub-system) by relying on the relevant discrepancy as a signal—authors disagree about the exact mechanisms, though they agree on these basic predictive workings (Floegel et al., 2023). We have an occurrence of metacognition according to the above thin notion of metacognition. This is unattractive; it implies that metacognition is involved in any predictive system of the mind (Carruthers, 2017). This is not the implication we wanted. We wanted a concept under which metacognition was more scarce.

One can respond to this challenge in different ways. Some authors seem sympathetic to the thin notion of metacognition and its implications that metacognition is very widespread in human and animal cognition (Shea, 2014; Shea & Frith, 2019). Other authors argue that although the idea of metacognitive feelings does involve a more liberal notion of metacognition, it does not carry the implication that predictive systems like the sensorimotor system involve metacognition (Proust, 2013).

These disagreements have shown that the dispute comes down to a difference between two basic positions on how to define metacognition. The dispute concerns whether metarepresentation is necessary for metacognition. Metarepresentation occurs when one representation represents another representation (Sperber, 2000). One of the philosophical positions claims that metarepresentation is not necessary for metacognition; i.e., some occurrences of metacognition do not involve metarepresentations. Defenders of this view are typically defenders of the idea of metacognitive feelings, as outlined above (Proust, 2013; Shea, in press; Arango-Muñoz, 2019). These authors claim that metacognitive feelings are non-representational. Such feelings do not represent the workings of the first-order system. Instead, they are non-representational affective signals that the meta-level system can use in its procedure of evaluating the first-order system. The view is known as proceduralism. The opposite philosophical position claims that metarepresentation is necessary to metacognition; i.e., all occurrences of metacognition involve metarepresentation. Defenders of this view are
typically critics of the idea of metacognitive feelings. They claim that metarepresentation must be a necessary element of metacognition, otherwise metacognition is too widespread in the operations of the mind, as stressed above. This position is known as metarepresentationalism (again, see Carruthers 2011, 2017).

The dispute between proceduralism and metarepresentationalism has been important to another line of philosophical discussion of metacognition. This line of discussion targets the role of metacognition perception-motor control. Perception-motor control concerns the interaction between an agent’s perception, motoric activity, and environment. It is a matter of an agent’s bodily action in her surrounding environment. Pacherie and Mylopoulos (2021) propose that athletic experts execute a kind of meta-control when they perform. This meta-control is a form of metacognition. Such meta-control does not concern the direct control of motoric activity or perception of the environment. Instead, it regulates the general modes of control that agents can execute motoric activity within. These authors distinguish between more cognitive and more automatic modes of control. Roughly, the first mode of control is slow and deliberate, while the second is fast and habitual. They claim that athletic experts shift competently and flexibly between these modes of control to secure optimal performance. This concerns both action selection (whether selection of goals is controlled or automatic) and action implementation (whether implementation of goals is controlled or automatic). This shifting is a form of control in itself, namely the relevant kind of meta-control. Such meta-control involves multiple metacognitive dynamics. The shifting between the modes of control involves a metacognitive evaluation in the form of a cost-benefit analysis. The cognitive control mode is more costly in terms of mental resources than the automatic control mode. The metacognitive evaluation computes whether the cognitive control mode is worth the cost considering the relevant benefits, or whether the automatic control mode is more appropriate (see also Mylopoulos, 2023). Shepherd and Mylopoulos (2021) make a similar proposal in discussing the skilled action of typists. They suggest that typists might draw on metacognitive feelings in monitoring their fine-grained sensorimotor activity. These feelings inform the typist about the unfolding of unconscious processing. Aligning with the ideas of general proceduralism, as discussed above (Proust 2013), such feelings are important to decision-making and the broader dynamics of control. Feelings of fluency might indicate success (e.g., suggesting that automatic control modes are sufficient), while opposite feelings might indicate problems in processing (e.g., suggesting that cognitive control modes might be needed).

The debates on proceduralism, metarepresentationalism, and metacognition in perception-motor control are important for the present PhD thesis. They constitute the most
sophisticated and well-developed discussions on the nature of metacognition in the philosophical literature. Hence, they serve as an academic background for the present thesis. Nonetheless, I would also like to stress that Articles 1–4 do not aim to directly tackle the dispute between proceduralism and metarepresentationalism. Moreover, Articles 1–4 make no commitment to any of these positions; neither do they argue in favor of any one of them. The articles are compatible with both views. Instead, the purpose of Articles 1–4 is to provide novel analyses of the functions of metacognition in human cognition that neither defenders of proceduralism nor metarepresentationalism have considered. These analyses focus upon metacognition in attention control and execution of skill, particularly as involved in mindfulness and flow. These kinds of metacognition are importantly different from the ones reviewed above. The above philosophical discussions typically approach metacognition as a sub-personal mental process. The metacognitive processes that produce metacognitive feelings are subconscious processes of cognitive sub-systems. Even though the subconscious processing of these sub-systems gives rise to conscious signals (in the form of feelings), the processing itself is not conscious or something we can ascribe to the agent as a whole. Agents are merely passive in producing the metacognitive feeling, although they receive the feeling and can implement it in further personal-level processing. In contrast, Articles 1–4 consider the dynamics of metacognition in mindfulness and flow. These are personal-level mental processes that are conscious and where agents seem to play a more active role in the metacognitive process itself. Let me review the relevant literature on mindfulness, flow, and their involvement of such metacognition.

2.2 Mindfulness and flow
Mindfulness research is a large scientific and academic field (Baminiwatta & Solangaarachchi, 2021). It involves disciplines such as philosophy, cognitive psychology, psychiatry, neuroscience, and Buddhist studies. Originating in Buddhist practice, mindfulness is today a standard element of the most efficient cognitive psychotherapies known as third-wave cognitive-behavior therapies (Hayes et al., 2011; Dimidjian et al., 2016). Researchers distinguish between two core meditation exercises of mindfulness. First, focused attention meditation involves subjects deliberatively focusing and sustaining their attention on a given object (like their own breath). They bring their attention back to this object if it has wandered. Second, open monitoring meditation involves subjects introspecting their own occurring mental states and relating to these with certain attitudes, notably the attitudes of
disidentification or nonreactivity. Many philosophers, psychologists, and cognitive neuroscientists stress that these exercises of mindfulness involve a form of metacognition at their core. In basic terms, they involve agents having explicit awareness and regulation of their own mental processes.

In the philosophical literature, Zawdiski (2019) proposes that mindfulness, and meditation more broadly, involves the exercise of a certain metacognitive skill. This skill concerns the capacity of agents to regulate their own mental states, in particular their own emotions, such that they are beneficial to the agent’s well-being. Relatedly, Leder and Zawidski (2023) propose that mental illness and disorder rest upon a dysfunction of this capacity. They argue that individuals suffering from disorders or illnesses like various forms of depression or anxiety are incapable of skillfully regulating their mental states and, notably, controlling their occurring pathological thoughts and feelings. Repetti (2018) has provided related analyses of mindfulness and meditation. Repetti proposes that mindfulness trains metacognitive awareness and metacognitive regulation of one’s own mental states. Such metacognitive awareness and regulation enables agents to control the mental states that govern their downstream processing. It enables them to be conscious of the mental states they undergo and, if relevant, hinder mental states in determining further processing and behavior. For example, agents might hinder harmful desires, such as those involved in addiction, in determining behavior (see also Latham, 2016).

The empirical and clinical literature on mindfulness aligns with these philosophical ideas. Multiple researchers argue that mindfulness is essentially a form of metacognition. Jankowski and Holas (2014) propose that mindfulness involves various metacognitive functions, such as metacognitive knowledge of how to practice meditation, metacognitive skills in regulating attention, and metacognitive experiences of having awareness of one’s own mental life. Dorjee (2016) proposes that the great diversity of different contemplative practices, including mindfulness, are unified by training the same metacognitive capacity for self-regulating one’s own mental processes and associated states of consciousness. At the center of this capacity is the ability of attention control. On a similar note, Garland et al. (2010) argue that mindfulness is a metacognitive form of attention control, since it involves ongoing monitoring of focus, emotions, affects, thoughts, and other mental processes. This tendency to view mindfulness as

---

3 Notice that Article 3 provides a more detailed introduction to mindfulness research. It also discusses the aspect of compassion in mindfulness, which I leave out above.
a metacognitive phenomenon is the dominant approach in the contemporary science of mindfulness.

Both these philosophical and psychological theories of the metacognition in mindfulness view metacognition as a personal-level mental process. It is a conscious mental process we can ascribe to the agent as a whole. The metacognition of monitoring and regulating one’s own mental life in a mindful way is something that the agent actively does. This is in contrast to the preceding philosophical discussions of metacognition in epistemic functions or perception-motor control. Let me outline how Articles 1–3 aim to contribute through novel analyses of the metacognition of mindfulness.

Article 1 provides a novel account of so-called decentering. Decentering concerns the ability of agents to move out of immersion in their own mental states. This ability is at the core of mindfulness, and authors understand it as a form of metacognition. Many clinical psychologists argue that decentering is one of the most, if not the most, important aspects of contemporary third-wave cognitive-behavior therapies (Bernstein et al., 2015, 2019). However, decentering remains unknown to most authors in philosophy. Article 1 provides a novel analysis of decentering as a complex mental operation composed of two sub-operations, namely introspection and detachment. This analysis should be of interest to both those familiar and those unfamiliar with decentering. It is a novel account that provides new insights into decentering, drawing on well-known philosophical ideas concerning attention, consciousness, and agency.

Article 2 discusses the Transparency thesis known from the philosophy of consciousness and perception. The article argues that mindfulness involves certain metacognitive dynamics that establish a form of introspective awareness that is in conflict with the Transparency thesis. This argument stresses an aspect of metacognition that previous philosophical discussions—such as the dispute between proceduralism and metarepresentationalism—do not commonly target. That is, Article 2 considers how metacognition (in the form of mindfulness) can induce radical phenomenological changes to how we undergo experiences. In the present case, such metacognition alters our experiences from being transparent to being opaque. This is a fundamental change.

Article 3 focuses on the idea that mindfulness is a form of metacognition—more precisely, a form of metacognitive control. As stressed above, this idea is widespread both among philosophers and psychologists. Article 3 offers two contributions. First, the article presents and defends the negative claim that current proposals concerning how to conceptualize the metacognitive control of mindfulness all fail. Relevant philosophers and psychologists alike
seem unaware of this unfortunate situation. This makes the negative claim important, since it stresses a basic issue in the dominant approach to understanding mindfulness in scientific terms. Second, the article offers a novel positive account of how to conceptualize the metacognitive control of mindfulness. The article argues that we should understand this metacognitive control as the implementation of metacognitive goals. To my knowledge, the explicit notion of metacognitive goals is new to both philosophical discussions and the psychological research of metacognition. Yet, it offers a fruitful approach to understanding mindfulness in scientific terms.

In conjunction, Articles 1–3 stress that agents can execute certain kinds of metacognition that acutely alter their state of consciousness, notably in terms of affects, emotions, and mood. The philosophy of metacognition should acknowledge this role of metacognition.

Turning to Article 4, this one aims to contribute to the growing literature on metacognition in skilled action and athletic expertise. The article considers the phenomenon of flow. Flow is a psychological state where agents execute their activity with focus, fluency, and a feeling of enjoyment or meaning. Many psychologists and philosophers have argued that flow is the characteristic state of athletic expertise. Among these, philosophers of the so-called Automatic view have made this argument (the below §4 on the philosophy of action explicates the core claims of this view). Contrary to common views on flow in philosophy, Article 4 argues that flow states involve multiple forms of conscious processing, among these forms of metacognition.

Article 4 draws on empirical work in sports psychology to support this claim. The psychological literature on metacognition in perception-motor expertise contains numerous proposals regarding the role of metacognition in athletic expertise and skilled action. This includes proposals on how metacognition as a personal-level mental process plays an important role. Brick et al. (2016) found that elite runners deliberated on attention strategies before performance (i.e., strategies for how to deploy their attention in certain situations). While performing, these runners explicitly regulated their attention in according with these strategies. Multiple studies in sports psychology also suggest that metacognitive dynamics (as those of mindfulness) are and can be of crucial importance to athletes in managing performance pressure. Such metacognition enables athletes to downregulate the effects of anxiety on their perception-motor execution and concentration on the relevant task (MacIntyre et al., 2014). Moreover, sports psychologists generally stress the importance of motivational self-talk among athletic experts. Such self-talk is a form of self-directed inner speech through which athletes motivate themselves. This motivation concerns sustaining focus of attention and generating
emotions and thoughts fruitful for performance. Studies suggest motivational self-talk to be a form of metacognition. Athletes report that they talk to their own heads—i.e., they aim mainly to regulate their own mental processes as attention, motivation, emotions, cognitive construal of the situation, etc.—when they engage in this kind of inner speech (Bellomo et al., 2020; Brick et al., 2020). Article 4 draws on this psychological literature in arguing that metacognition is involved in perception-motor expertise as a personal-level mental operation, also in the states of flow. This is an alternative approach to metacognition in perception-motor functioning compared to that of the philosophical research reviewed in §2.2 above.

3. Attention

For a long time, philosophical discussions of attention mainly concerned the relationship between attention and consciousness. One important discussion concerned the dispute on whether attention is necessary and/or sufficient for consciousness (Block, 1995). However, over the last 15 years or so, philosophers have begun to approach the topic of attention in alternative ways (Mole et al., 2011). One major focus has been on the role of attention in action. This focus involves the relationships between attention and agential aspects such as goals, control, perception, and skill. Let me outline how Articles 1–4 fit into these discussions.

3.1 General approaches to attention

Articles 1–4 approach attention as a personal-level phenomenon. This approach is mainstream in current philosophical psychology. Many philosophers think that approaching attention as a personal-level phenomenon enables us to tackle a kind of eliminativism about attention (expressed by numerous cognitive scientists: See Hommel et al., 2019; Anderson, 2011, 2021; Taylor, 2023; Watzl, 2023, for philosophical discussions.). This eliminativism claims that we should avoid using the concept of attention in scientific practice, since it has no consistent and unifying meaning across various research branches. Further, eliminativists often claim that the various functions we ascribe to attention (e.g., workings in different perceptual modalities, various cognitive functions as in reasoning or planning, motor orientation) do not involve a distinct and clear neural implementation (the neural processes of what we call the workings of attention are vastly spread all over the brain). Hence, it does not make sense to talk about attention in a scientific sense. Many philosophers have claimed that such eliminativism
assumes attention to be a sub-personal processes, and that this is a mistake. Instead, we should consider attention exactly as a personal-level mental process (Watzl, 2017, 2023; Wu, 2016; Dicey-Jennings, 2020; Mole et al., 2011). We should not expect to account for attention as a distinct sub-system of the person with a clear and distinct neural implementation. Attention is a mental process we can ascribe to the person as a whole, and it might involve several different neural workings in various contexts. The job of a philosophical theory of attention is to provide an account of attention that can adequately describe and explain these multiple workings in a unifying way. Let me review two philosophical theories of attention that have been central to the writing of Articles 1–4.4

First, consider the structure theory of attention (henceforth the STR). STR claims that attention is not a specific kind of mental state, but instead an organization of mental states into a priority structure of the agent (Watzl, 2017, ch. 5). Priority structures are arrangements of mental states into the fore- and background of an agent’s mind. The foreground means two things. One, the more in the foreground a mental state is, the more decisive it is in determining an agent’s state of consciousness or the content of her conscious representation (p. 4, p. 147). Two, the more in the foreground a mental state is, the more prominent the state is in determining the processing of the agent (p. 107, p. 151). Conscious representation and determination of processing are thus associated in STR’s notion of priority structures. Moving a mental state to the foreground of an agent’s priority structure involves boosting both its conscious representation and its causal potency in determining further processing. STR views attention control as an active control of one’s own priority structure. Individuals execute attention control when they structure or restructure their priority structures in accordance with their relevant intentions or goals (p. 142). Essentially, STR views attention as an ordering aspect of the mind that prioritizes certain states above others.

Second, consider then the selection for action theory of attention (henceforth SfA), as developed and defended by Wu (2011, 2016). SfA builds upon work in cognitive psychology, such as work by Neumann (1987) and Allport (1987), certain operationalizations of attention in experimental psychology, and the biased competition model of attention mentioned above. The many-many problem is a central idea in SfA. Wu (2011, 2016) claims that this problem is a fundamental situation of agency. The problem concerns the case that agents always receive multiple inputs that they can couple with multiple outputs. Inputs can be perceptual inputs (as

---

4 One way to understand this task is that a philosophical theory of attention must make sense of our pre-theoretic notion of the nature of attention. This notion states that attention is the “selective directedness” of our minds (Mole, 2021). The job of a philosophical theory is to illuminate the nature of such directedness.
in visual or auditory perceptions) or more cognitive inputs (as in thoughts). Outputs can be both bodily movement (as in raising one’s arm) or mental operations (as in further reasoning). This landscape of multiple inputs and outputs constitutes a problem insofar that to do anything an agent must select one of the many inputs and couple it with one of the many possible outputs. To act, an agent must select an input–output coupling where the output is either a bodily or a mental action. SfA claims that attention is the selection of an input for an action. In bodily action, the action involves overt bodily movement; in mental action, the action involves covert mental activity. Building upon this account, SfA claims that the nature of attention control is that of structuring this selection of an input for an action. An agent controls her attention when her relevant goals or intentions structure the selection of an input for some action, such as when an agent goal of finding a coffee shop structures her selection of a visual input of a coffee sign with the motoric output of moving toward this sign.

STR and the SfA do not exhaust the supply of philosophical positions. Many other authors have developed and defended alternative theories. Dicey-Jennings (2020) argues that attention is the personal-level mental selection of objects (without this selection necessarily involving action), while Fazekas and Nanay (2021) argue that attention is not to be thought of as selecting some object but as amplifying an input signal in the cognitive system. Moreover, Mole (2011) argues that attention is not a type of personal-level psychological process, even though it is a kind of personal-level mental phenomenon. Mole claims that attention concerns the way in which psychological processes work together and unfold when processing a task. An agent can process a task attentively or inattentively. This comes down to the degree of cognitive unison in her processing. Such unison is the degree to which the agent deploys her relevant cognitive capacities in processing the relevant task (instead of these capacities processing other information or not being active).

Despite these differences, philosophers and psychologists agree on a set of interrelated distinctions concerning attention (Chun et al., 2011). Voluntary versus involuntary attention concerns the difference between instances of attention that agents control and those that they do not. Goal-driven versus stimulus-driven attention concerns the difference between instances of attention that are driven by an agent’s relevant goals and those where it is driven by an agent’s exposure to acute stimuli. Endogenous versus exogenous attention concerns the difference between instances of attention that are activated by internal states of an agent and those that are activated by external objects. Lastly, internal versus external attention concerns the difference between instances of attention that involve mental states that do not represent an acutely present external object (e.g., thoughts about what to have for dinner or planning about
where to travel this summer) in contrast to instances that involve mental states that do represent acutely present objects (e.g., one’s visual state of seeing the cup at the table or hearing the sirens passing the window). Historically, many authors have grouped these distinctions into two basic realms of attention (Watzl, 2011a, 2011b). They have considered voluntary, goal-driven, and endogenous attention as constituting one realm of attention, while involuntary, stimulus-driven, and exogenous attention constitute another. However, the workings of attention that Articles 1–4 investigate challenge this grouping. These articles place a specific focus on the attention involved in rumination (which is, roughly, past-oriented repetitive thinking about negative events) and worry (which is, roughly, future-oriented repetitive thinking about possible negative events). Clinical and sports psychologists view rumination and worry as patterns of attention (Wells & Matthews, 2014). As such attention patterns, rumination and worry seem to combine these distinctions of attention in interesting ways. Rumination and worry are involuntary in the sense that these states occur without agents intending them to do so. Further, as stressed above, states of rumination and worry are also counterproductive to an agent’s relevant goals. Rumination and worry are considered malfunctioning patterns of attention that disrupt mental health or the focus needed for performing. The depressive patient’s rumination is counterproductive to her goal of feeling well, as the worry of the athlete is counterproductive to her goal of performing well. Yet, many instances of rumination and worry are endogenously generated. These patterns do not need direct environmental stimuli or triggering but occur in resting states or from “within” (Wells, 2011). Lastly, rumination and worry are usually internal forms of attention; i.e., they do not involve mental states that represent acutely present external objects. Instead, the states of rumination and worry represent more distant events or properties of these (negative aspects of past or future situations). Consequently, it seems clear that rumination and worry are special manifestations of attention. They are involuntary, endogenously generated, internal occurrences of attention that are not directly goal-driven but are counterproductive to an agent’s relevant goals (see also Watzl, 2011a, 2011b, 2017, for similar points on mindwandering). Articles 1–4 examine how agents control such special workings of attention.

3.2 Introspective attention
Let me then shift to the philosophical discussion of introspective attention and what states of awareness such attention can give rise to. Many philosophical discussions consider introspection from an epistemological perspective; i.e., introspection is the process of
accurately gaining self-knowledge about one’s own mental states (Schwitzgebel, 2019). However, philosophers disagree about how such introspection works. One basic disagreement concerns whether introspection is a matter of turning one’s attention inward. Some philosophers deny this (e.g., Evans, 1982). Yet, most philosophers seem to agree that many forms of introspective processes involve some kind of inward turn in attention (though they disagree about the exact workings of such inward attention: Carruthers, 2011). The present PhD thesis also considers introspective workings that involve some kind of inward attention. Specifically, the thesis targets two topics concerning detachment and transparency in introspective attention. Let me outline these separately.

Article 1 concerns the relationship between introspective attention and detachment. Typically, many philosophers and cognitive scientists stress that introspection of a mental state makes the state more conscious. This idea seems present in both Watzl’s (2017) development of the STR and Wu’s (2011, 2016) formulation of the SfA. However, it is not obvious exactly what this idea involves. One way to understand it might be to follow James (1890) in thinking that consciousness is organized into more focalized and periphery areas, such that objects or states in the focalized area are more conscious in the sense that they are more vivid or stronger when experienced by the subject. Introspecting a mental state is a matter of focalizing it in this sense (notice that both Watzl, 2017, and Wu, 2011, 2016, explicitly draw on James, 1890).

I shall not discuss whether this idea is plausible. For now, the important thing is that many authors believe that consciousness serves two functional roles in cognitive systems. First, unlike nonconscious mental states, conscious mental states enable systems to integrate various information sources, from different cognitive sub-systems into broader multi-modal representations. Second, conscious mental states are distributed globally across various cognitive sub-systems, unlike nonconscious states that are not distributed globally and are accessible across various sub-systems (Dennett, 1991; Dehaene & Naccache, 2001). Conscious mental states thus enjoy a certain prominence. They are multi-modal integrated representations globally available to various sub-systems, meaning that these states are highly determinative of an agent’s processing and associated behavior. Putting these considerations together, one can draw some connections: When an agent introspects a mental state, this makes the state

---

5 Notice that this claim is different from claiming that introspection is necessary for consciousness, i.e., making the claim that introspection of a mental state is necessary for that mental state to be conscious. This is, obviously, a controversial claim relating to discussions of higher-order theories of consciousness (Lau & Rosenthal, 2011).
more conscious, meaning it also becomes more determinative of the agent’s processing (again, both Watzl, 2017, and Wu, 2016, seemingly endorse such an idea).

Article 1 explores this idea of introspection by focusing on the operation of decentering. An agent decenters from a mental state when she introspects and detaches from it at the same time. This stresses a role of introspection that is not commonly discussed in philosophy. First, it stresses that introspection is involved in making states less determinative of processing (opposing the above considerations). Second, as stressed above, philosophers usually portray the purpose of introspection as that of gaining self-knowledge about one’s own mental states. In the case of decentering, the role of introspection is more one of aiding psychological control of one’s own mental processes.

Article 2 targets another well-known discussion in philosophy, namely the discussion of the Transparency thesis (from now on TT). This thesis is prominent in philosophy of consciousness and perception. Authors outline it in slightly different ways. One interpretation is phenomenological. This interpretation understands the TT as a phenomenological thesis excluding certain states of introspective awareness (see discussions in Kind, 2003; Mitchell 2020). The TT claims that when an agent is introspecting her experience, only objects and properties of the objects of her experience can appear to her. That is, when an agent introspects her experience, she can only be aware of the objects and properties of these objects that her experience involves. Another interpretation is more metaphysical. This interpretation does not consider the TT as a claim about how something appears to subjects. Instead, the metaphysical interpretation views the TT as a thesis on how to account for the metaphysical nature of the properties that appear in introspective awareness (see, for example, Weksler et al., 2019).

One could think that there is an important difference between these two interpretations. They might have different evidential relations, such that disconfirmation does not transfer between them. The phenomenological version of the TT is disconfirmed by cases where agents have introspective awareness of features that genuinely do not appear as objects or properties of objects of their experience. Such cases might not directly disconfirm the metaphysical version of the TT. The metaphysical version of the TT is disconfirmed by cases in which agents are aware of some feature(s) that we cannot make sense of as objects or properties of objects of their experience. On the other hand, if we can point to features in introspective awareness that we cannot make sense of as objects or properties of objects (hereby disconfirming the metaphysical version of the TT), then this does not necessarily disconfirm the phenomenological version of the TT. It depends upon whether these features genuinely appear to agents as objects or properties of objects. They need not do so. Philosophically, we might
succeed in arguing that some properties cannot be accounted for as objects or properties of objects, yet these properties still appear to be so in the introspective awareness of subjects.

Article 2 considers the phenomenological interpretation of the transparency thesis. I refer to this interpretation when I write “TT” henceforth. Proponents often present and defend the TT as a general thesis covering all modalities of experience and introspective awareness. There is a good reason for doing this. The TT has been an important element in broader metaphysical theories of consciousness, notably Representationalism (as formulated in Tye, 2002; Harman, 1990; Dretske, 1997). Roughly, Representationalism claims that we can adequately understand consciousness, or experience, as a form of representation. Consciousness is a way that our minds can represent objects. Many philosophers have argued that this view is the most promising in providing a naturalistic theory of consciousness. Defenders of Representationalism have provided various arguments, yet the claimed truth of the TT has been among the most central. It makes good sense: If we cannot become introspectively aware of anything other than appearing objects of our experience and properties of these, then there are plausibly not any other properties involved in our experience. That is, consciousness is nothing more than a representation of objects and their properties. This is how the phenomenological version of the TT would confirm a metaphysical theory as Representationalism.

The TT has been subject to manifold criticism. However, Article 2 develops and defends a novel criticism of the TT. This is the criticism that mindfulness enables states of introspective attention and awareness that oppose the TT. This idea is widespread among philosophers and psychologists researching mindfulness. However, the idea has remained unclear in various crucial respects, making it difficult to discuss in precise philosophical terms. Article 2 provides an elaboration of the idea and labels it the Mindfulness opacity hypothesis (from now on MOH). Opacity and transparency are mutually exclusive in the sense that when our experiences are opaque under introspection, then we are aware of features that genuinely do not appear as objects or properties of objects of our experience. Article 2 draws on both philosophical and empirical research in developing the MOH. The article concludes that the MOH is well motivated.
4. Action

The philosophy of action aims to describe and explain the nature of action. Traditionally, one might say that this literature targets the “problem of action” (Davidson, 1971). This is the problem of answering the question: What separates actions from mere happenings? The present PhD thesis assumes a certain theoretical approach to this matter. Articles 1–4 work under the group of theories labeled causal theories of action, broadly understood (for a good introduction, see Glasscock & Tenenbaum, 2023). These theories aim to account for the nature of action by its characteristic causal workings (typically stressing how certain mental states or attitudes play the role of causal antecedents or guiding factors in action). Relatedly, this group of theories approach action in naturalistic terms, meaning that they view the philosophical study of action as continuous with the relevant empirical and scientific investigations of the topic. Nevertheless, causal theories famously disagree about the more exact causal workings of action (REF). This disagreement is not important to the present PhD thesis. The thesis does not assume commitment to any specific causal theory of action; nor does it argue in favor of a certain position. Instead, Articles 1–4 target two more specialized discussions in the current philosophy of action, namely the discussions of skilled action and mental action. Let me review these separately.

2.1 Skilled action

One major topic in recent philosophy of skill and expertise has been the issue of how controlled and automatic processes interact in the execution of perception-motor expertise (here, perception-motor expertise refers to the kind of skills and expertise involved in athletic domains). The differences between controlled and automatic processes are a point of disagreement, both conceptually and empirically (see Fridland, 2017; Wu, 2016). However, many authors agree that automatized processes are characterized, at least to some degree, by four basic features (Bargh, 1994):

(i) They are initiated without conscious intentions of the agent.
(ii) They occur outside of an agent’s awareness.
(iii) They cannot be vetoed; i.e., an agent cannot cancel them.
(iv) They are efficient in the sense that they do not draw upon limited cognitive resources.
The features of (i)–(iv) provide a negative characterization of automatic processes. They characterize automatic processes by their lack of conscious intentions, awareness, veto-control, and conscious attention. It is standard to define controlled processes as the opposite, namely as the processes that involve conscious intentions, awareness, possibility of vetoing, and conscious attention. Historically, many researchers have described the relationship between automatic and controlled processes as dichotomous. They describe controlled and automatic processes as opposites, hence as types of processes that exclude each other in action control. According to this dichotomous view, either an action is controlled or it is automatic (Wu, 2016).

However, today, many philosophers and psychologists dispute this dichotomous idea of controlled and automatic processes. First, they claim that whether a process is controlled or automatic is not a categorical property but a gradual or scalar property (i.e., processes are more or less automatized). Second, and more importantly to the present PhD thesis, authors argue that controlled and automatic processes are complementary in the sense that they most often interact in action, meaning that they do not exclude each other. The idea of control hierarchies is central here. Multiple authors argue that action control is composed of a hierarchy of goals (Pacherie, 2008; Christensen et al., 2016; Shepherd, 2015, 2019). These multiple levels of goals interact in action control. Take the example of an agent riding back to her apartment on her bike. This involves the more ultimate goal of getting to the apartment; it involves the more proximal goal of navigating the streets; and it involves the motor goals of making fine-grained motoric movements. Philosophers generally think that the goals at the higher end of the hierarchy (the ultimate and some proximal goals) are associated with controlled processes. Selection and implementation of these goals involve conscious intention, awareness, and ability of vetoing, and they demand resources of conscious attention. Goals further down in the hierarchy are constrained by goals at the higher level. However, these lower-level goals are associated with more automatic processing (e.g., the fine-grained motoric activity of riding your bike—changing gears, pedaling, balancing, etc.—unfolds without the need of your conscious intentions or awareness; it is so quick you cannot veto it in many cases; and it has no need of resources of conscious attention). Nevertheless, these are controversial matters. Some authors argue that conscious attention and guiding are involved all the way down to the very low and fine-grained levels of motor goals (Fridland, 2015), while other authors argue that agents only control the hierarchies at the higher level in selection of goals and not in implementation of these at the lower levels (Papineau, 2013, 2015). Here, the important thing is that the overall idea of control hierarchies provides an argument against the dichotomous
view of automatic and controlled processes. These hierarchies show that controlled and automatic processes interact in action control, not that they exclude each other (Pacherie, 2008; Shepherd, 2015, 2019; Wu, 2016).

The general idea of control hierarchies is of relevance to a more specialized discussion of optimal athletic expert performance in philosophy and psychology. Optimal athletic expert performance refers to the situation in which elite athletes perform well. Philosophers have been interested in these situations, since they provide a window into agency in general. If we can understand the psychological and bodily workings of optimal athletic performance, we might be able to gain insight into the core nature of motoric skill and expertise. For many years, the literature was dominated by a specific position on this issue, the so-called Automatic view of expertise. I refer to this view as the AV from now on (for proponents of the AV, see Dreyfus, 2007a, 2007b). We can outline the central claim of the AV as follows: When athletic experts perform optimally, their automatic processes run freely without conscious supervision. Here, conscious supervision mainly refers to three interrelated aspects, namely conscious control (i.e., conscious selection and implementation of goals), self-awareness of performance (i.e., self-awareness of the success of one’s activity), and conscious task-related thoughts (i.e., thoughts about one’s own activity, as in strategic or evaluative thoughts). The AV claims that optimal athletic expert performance is without such elements of supervision. For example, an expert tennis player does not consciously supervise the perceptual dynamics (e.g., her visual attention shifts) or motoric workings (e.g., positioning of the arm) that constitute her expert level backhand. Given that she performs optimally, these processes are simply automatic processes that run freely, without the athlete monitoring or regulating them consciously as they unfold.

Criticism of the AV has been a major thread in contemporary philosophy of skill and expertise (Pacherie & Mylopoulos, 2021; Fridland, 2017; Christensen et al., 2016). Yet, this criticism has not targeted a very specific argument for the AV, namely the argument on so-called flow states. The idea of flow was introduced to scientific psychology by Csikszentmihalyi (1975). The exact conceptualization of flow states is a topic of controversy in current psychology. However, I think the most adequate and uncontroversial conceptualization is to say that an agent is in flow when she executes her activity with fluency, deep focus, and with a sense of enjoyment or meaning to it. The research literature on flow states is extensive (Abuhamdeh, 2020). It contains many disputes relating exactly to conceptualizations of this type of state, but also to measurements of it (Swann et al., 2018). Here, the important thing is that a key argument of the AV concerns flow states (Dreyfus,
In Article 4, I propose that we label this the flow argument (FA). Article 4 outlines the FA as follows:

(1) Flow is the characteristic state of optimal expert performance.*
(2) If an agent is in the state of flow, then (a) she does not consciously control her actions, (b) she is not self-aware about what she does, and (c) she does not have conscious task-related thoughts.
(3) If an expert performs optimally, then she is in a state characterized by claims (a)–(c).

The FA has been highly influential in both philosophy and psychology. Multiple philosophers, who do not otherwise support the AV, seemingly endorse the FA or central parts of it (Annas, 2008; Velleman, 2008; Railton, 2009). Likewise, many psychologists accept this account of flow and expert performance, both in the academic and more popular literature (Duckworth, 2007).

Despite this prominence and importance of the FA, the argument remains relatively unexamined. This is unfortunate, since the success of the FA is important to evaluating the AV as a broader position. Moreover, if the FA is mistaken in its portrayal of flow, then this suggests that there are important misunderstandings of flow in the broader philosophical and psychological literature. Article 4 aims to provide an in-depth examination of the FA. The article argues that both premises of the FA are implausible in the light of current psychological research on flow. The article also draws on plausible philosophical background theories, in particular relating to control hierarchies, in making this claim. One core idea of Article 4 is that personal-level mental processes, in the form of conscious supervision, do appear under flow states. Such processes are present in various ways, and they positively facilitate the flow states. More precisely, Article 4 argues that agents consciously control their attention under flow states; that psychological mindsets, like mental toughness, are active under flow states and that this involves conscious thoughts and cognitive interpretations of events under performance; and that athletes utilize motivational self-talk and metacognitive attention strategies under flow, which are mental processes associated with conscious inner speech and monitoring of one’s own attentional focus.

---

6 The ‘*’ indicates that we can give multiple interpretations to premise (1). Article 4 interprets it as a conditional, meaning that the argument as a whole is interpreted as a transitive argument.
2.2 Mental action

Another major trend in contemporary philosophy of action is the topic of mental action. Mental action concerns the issue of whether and why some mental processes of an agent qualify as actions of that agent while other mental processes do not. This makes the philosophical debates on mental action similar to the problem of action referred to above (the problem of why some bodily movements qualify as actions and other are mere happenings). Some authors express general skepticism of the idea of mental action, arguing that the idea is incoherent in different ways (Strawson, 2003; see also Levy, 2019). Nonetheless, such skepticism seems to be a minority position today. Most philosophers now endorse the idea of mental action and consider it in various respects. Despite this widespread interest, there is currently no standard definition of mental action that states the requirements for such action. However, I expect that many philosophers would be sympathetic to the below definition, at least as a rough sketch.7

A mental process, M, of an agent, A, qualifies as mental action of A if and only if 
(a) M is intentional to A under some description.
(b) A controls M.
(c) M is covert.

The issue with this definition is that each condition is underspecified as it stands above. Consider, for example, condition (b). Philosophers have different views on how agents control their bodily actions (Shepherd, 2014; Buehler, 2022; Wu, 2016). These disagreements plausibly transfer to issues on control in mental action; hence, the requirements for satisfying (b) might be controversial. Consider also condition (c). It is not obvious what this condition demands exactly. Does it demand that mental actions do not involve any overt behavior at all (e.g., no eye movements or motoric activity)? This seems too strong. As we shall see below, many authors think that attention shifts in perceptual attention involve mental actions. That is, when agents make shifts in visual or auditory orientation in such attention, they execute a basic mental action, even though these mental actions involve overt behavior in the form of eye movement or head movement. Hence, one might interpret condition (c) differently. One might think that the condition demands that no overt behavior is essential or exhaustive to mental actions. For example, if an agent taps her finger on the table while reasoning, we can still claim that her reasoning is a mental action because this tapping is a nonessential aspect of the mental

---

7 I draw on Tollerup and Grünbaum (in preparation) here.
action (i.e., the mental action could occur without this movement), or because we think that the tapping of the finger does not exhaust the mental action (i.e., reasoning is not merely a matter of this overt behavior). Likewise, shifts in perceptual attention, involving eye or head movements, would still be examples of mental actions because agents could execute these attention shifts without these movements (i.e., without moving their eyes or head), or because such overt behavior does not exhaust what the agent does under such shifts (there is something more to these shifts than the movement).

Even though these issues are important, and future philosophical research should address them, I consider the above definition sufficient as a background for the discussions of mental action in the present PhD thesis. The cases of mental action that this thesis discusses should be clear-cut examples of mental action under various theories of control and interpretations of covertness. That is, Articles 1–4 consider mental actions in the form of personal-level mental operations involved in the metacognition of mindfulness or the psychological aspects of flow experiences (such as self-talk or mindsets that are active under performance).

Traditionally, philosophers have been interested in mental actions in two interrelated respects, namely the role of mental action in various epistemic functions and the role of mental action in attention control. Consider the first one. Multiple authors argue that core epistemic functions such as decision-making (Proust, 2013) or reasoning (Sorotiou, 2013) are paradigmatic examples of mental actions. These mental processes satisfy conditions (a)–(c) above. The present thesis is not an attempt to contribute to these current discussions.

Then, consider the second focus, concerning attention control. Multiple philosophers claim that attention control is a basic form of mental action. As mentioned above, these authors consider shifts in attention or the active sustaining of attention as forms of mental action (Buehler, 2022; Watzl, 2017; Wu, 2016). Depending upon which general philosophical theory of attention one commits to, one can flesh out this idea in various ways. Recall the structure theory of attention (STR). STR claims that attention is a structure of the mind, i.e., an ordering of mental states into a fore- and background (Watzl, 2017, 2022). According to STR, attention control is the activity of organizing this structure in accordance with one’s relevant goals or intentions. This elaborates on the present idea: The mental action of attention control is the action of moving certain mental states in the foreground of one’s own priority structures. Then, recall the selection-for-action theory (SfA). SfA stresses that attention control is a matter of structuring the selection of an input for action by one’s relevant goals or intentions. The mental action of attention control would be the capacity of changing this structuring or, in other words, changing the selection processes by which inputs are coupled with outputs. This involves
modification of the weight distribution of the relevant landscape of input–output couplings (Wu, 2016). Here, the important thing is that philosophers tend to view these mental actions of attention control in relation to perception-motor activity or in epistemic endeavors. By this, I mean that they tend to stress the importance of such mental actions in relation to agents’ bodily movements, representations, and endeavors to achieve goals in the external environment.

Articles 1–3 propose another function of the mental action of attention control. These articles consider the mental action of attention control in terms of regulating one’s own affective, emotional, and mood-related states. This perspective is not completely new to the philosophical literature. Other authors—notably in the philosophy of meditation—have already argued that the mental action of attention control is important in emotion control and the management of distressful states like pain (Poletti et al., 2021). Nonetheless, Articles 1–3 aim to offer new accounts of these dynamics. Article 1 argues that the phenomenon of decentering, and hence mindfulness, involves agents establishing and maintaining certain stances toward their own occurring mental states. As stressed earlier, these are the stances of disidentification and nonreactivity. These stances constitute mental actions of attention control, since the stances downregulate the determination of pathological mental states on further processing. However, the underlying goal is not that of moving in the environment or representing it—instead, it is that of regulating one’s own affects, emotions, and moods.

As stressed in previous sections, Article 2 argues that mindfulness enables a form of introspective attention and awareness that opposes the Transparency thesis. The article argues that this kind of introspection is a form of skilled mental action. It is a form of mental action where agents can intentionally and substantially change their state of experience, going from a state of transparent to opaque introspective awareness of phenomenological properties. This mental action is not about movement or accurate representation of the external environment. Instead, it is about changing one’s mode of experience.

Lastly, Article 3 argues that the essential and demarcating psychological capacity of mindfulness is a form of metacognitive control. Such metacognitive control involves agents implementing certain kinds of goals, namely metacognitive goals. Article 3 argues that this metacognitive control satisfies the usual requirements of mental action, i.e., the abovementioned conditions (a)–(c). Article 3, hereby, offers a new account of both why mindfulness is a form of metacognitive control and why this metacognitive control is a form of mental action. Again, this concerns a mental action that often has an underlying focus on affects, emotions, and moods.
5. Methodology and interdisciplinary contributions

Let me end this introduction with some considerations on the methodology of the present thesis. The previous sections have highlighted how this thesis aims to contribute to current philosophical discussions. Yet, it also aims to make more interdisciplinary contributions, in particular with relevance to current debates in the psychological literature. Let me clarify this interdisciplinary aim by considering three general and profoundly interrelated ways in which I think philosophy, and its methodology, can offer relevant analyses to empirical disciplines of the mind. I focus on psychology in particular here.

First, philosophy can offer clarifications of concepts in psychology. Philosophers are trained in assessing the definitions of concepts, their application, and their relation to other concepts. Article 1 aims to clarify the concept of decentering. Despite its prominence in current clinical psychology, the concept is still unclear with respect to various aspects of introspection, detachment, and attention control. Article 2 aims to clarify the concept of opaque introspection, which is widely discussed in the psychology and cognitive science of mindfulness. This concept is also unclear in various respects when it concerns mindfulness. Article 3 aims to clarify the concept of metacognitive control, also prominent in mindfulness research. The article argues that the current definitions of the concept, in both the philosophical and empirical literatures, are unsuccessful. Lastly, Article 4 provides a clarification of the concept of flow. It argues that there are broad misunderstandings of the concept in both the philosophical and psychological literature. Such clarifications are also of relevance to psychological measurement procedures. The issue of “construct validity” concerns the worry whether a given measurement tool (e.g., a scale) really measures the psychological attribute it aims to measure (e.g., does our well-being scale really measure well-being in the relevant population?). Many authors claim that to build a measurement tool that can measure the desired attribute, it is necessary that one builds this tool on a clear and plausible characterization of the relevant attribute (e.g., if our psychological scale is to measure well-being, we must operate with a clear and plausible concept of well-being and build our scale upon this conception. For further in-depth philosophical discussion of construct validity, see: Lange and Grünbaum, 2023). Hence, clarifications of concepts like decentering, opaque introspection, metacognitive control, and flow are substantially important for our ability to measure these psychological phenomena.

Second, philosophy can offer clarifications of theoretical positions in psychology. Philosophers are trained in examining the core claims and ideas of positions. Such examination
often involves clarifying the exact points of disagreement between the relevant positions. Notice that disagreement between positions might be a matter of how to define a concept properly. For example, Article 1 compares two different positions on how we could make sense of the construct of decentering: one that views the operation as a simultaneous execution of both introspection and detachment and another that views the operation as one operation of merely bare introspective attention. The article explains the substantial differences between these two views. Likewise, Article 3 examines the current positions on how to define mindfulness as a kind of metacognitive control. These positions are prominent in both philosophy and psychology, yet they have not been compared in great detail. In cases like Article 1 and 3, the clarifications of concepts and theoretical positions are overlapping in the sense that positions disagree exactly about how to define a concept. Yet, clarification of positions can also concern explicating assumptions, commitments, and predictions of a certain position. For example, Article 2 clarifies the central claims and ideas of the popular idea that mindfulness enables opaque introspection. As stressed above, this position is widespread in the scientific literature on mindfulness, yet authors do not elaborate on what this position amounts to. The article offers such an account and aims at a coherent formulation of the position. Lastly, Article 4 provides an extended analysis of the Automatic view in the philosophy and psychology of skill and expertise. The article provides clarifications of the central aspects of this view, in particular in relation to flow. This includes clarifying the kind of predictions that the Automatic view would make about observations in psychological studies.

Third, philosophy can offer clarifications and examinations of the relevant evidence needed to assess competing positions. Philosophers are trained in reading across a range of literature and systematizing relevant findings. Article 1 discusses whether the phenomenon of decentering disconfirms the popular selection for action theory of attention. Article 2 provides the first systematic and philosophical examination of the evidence motivating the idea that mindfulness enables opaque introspection. This concerns both empirical studies and philosophical theorizing. Article 3 discusses the kind of evidence that would support the idea that mindfulness is a form of metacognitive control. Finally, Article 4 provides a systematic account of psychological studies that disconfirm the Automatic view’s argument on flow states.

These three lines of analyses are novel to current discussions in psychology. My aim is that they provide relevant insights to psychological research.
5.1 Clinical relevance

Given the above considerations, it might sound as if Articles 1–4 attempt to make interdisciplinary contributions solely in terms of basic research. This is not the case. In fact, I hope that the work of the present thesis can be of relevance to more practical endeavors as well. This concerns work in clinical psychology and psychotherapy in particular. As already mentioned, mindfulness is a core component in contemporary clinical and psychotherapeutic protocols. I hope that Articles 1–3 are of relevance to such protocols. Numerous clinical researchers stress that we need new theories and frameworks of mindfulness that can illuminate the psychological and more personal-level dynamics of this practice (Dorjee, 2016; Garland, 2010; see also philosophical discussions in Repetti, 2022). Such theories should illuminate the dynamics of attention, cognitive control, goal-implementation, consciousness, emotion regulation, and self-awareness-involved mindfulness. This informs us on how mindfulness involves an agent’s mental life on general, holistic, and broader levels. This is different from the neurobiological measures of mindfulness that describe the various neural patterns involved in this practice (Zeidan et al., 2014). Articles 1–3 attempt to offer this form of analysis.

Notice that this ambition concerning clinical relevance is not uncommon in contemporary philosophical psychology. Other philosophers have had similar clinical ambitions, although usually with respect to other topics. In particular, this thesis has been inspired by two other lines of philosophical work with such an ambition: the work of Pickard (2011, 2017) and Bortolotti (2016; Bortolotti & Sullivan-Bissett, 2018; Murphy-Hollies & Bortolotti, 2022). Let me clarify the relevant connections between these lines of work and the present thesis.

First, consider Pickard’s (2011, 2017) idea of “responsibility without blame” in clinical practice. Pickard argues that clinicians should take the stance of responsibility without blame in interacting with clients in psychiatric and social institutions. This stance ascribes responsibility to clients, yet it ascribes this responsibility without blaming the clients. Pickard motivates this proposal in various ways. She argues that empirical studies suggest that clients of various patient populations are actually able to control their pathological behaviors. For example, many subjects with substance use disorder seem able to postpone their use of a substance—moreover, many subjects with a personality disorder can intervene on their damaging social behavior. Even though this might not amount to a total form of control, it grounds ascription of responsibility to the clients (at least to some degree). Further, Pickard argues that not only is this ascription empirically warranted, it is also fruitful in terms of treatment. Approaching clients with the stance of responsibility without blame is the best way
to support the psychotherapeutic progression. It is the stance that fosters the desired change in service users.

Second, consider Bortolotti’s (2016; Bortolotti & Sullivan-Bissett, 2018) work on delusions in relation to psychiatric conditions like schizophrenia. Bortolotti seemingly takes an epistemological *agency-first* approach in analyzing delusional beliefs (see for example Murphy-Hollies & Bortolotti, 2022). She claims that this approach differs from more traditional epistemology. Traditional epistemology tends to evaluate belief-like mental states in isolation (evaluating whether a belief is rational in relation to questions of evidence, coherence, process of formation, etc.). The agency-first approach differs from this traditional epistemology, since it does not evaluate belief-like states in isolation. Instead, the agency-first approach examines belief-like states in broader relation to an agent’s overall epistemic projects. It examines whether a mental state is fruitful for an agent in succeeding in her broader epistemic practices. Bortolotti (2016) argues that many of the delusions we find in psychiatric conditions, like schizophrenia, are irrational from the perspective of traditional epistemology; yet, from the agency-first perspective they do bring benefits to the epistemic practices of the agent. For example, delusions about one’s own identity or personal history can foster a sense of meaningfulness to subjects’ lives. Such meaningfulness plausibly enables subjects to take up conversation with other individuals and expose themselves to other epistemic practices, having broader positive epistemic consequences. One might think that whether a client’s delusional belief is beneficial from an agency-first approach is important to how a therapist should assess and interact with this belief. If the belief is beneficial from this agency-first perspective, the therapist might judge that the belief does not call for psychotherapeutic intervention or critical assessment.

The present PhD thesis is inspired by Pickard’s and Bortolotti’s general aim of making academic philosophy relevant to clinical practice in a very direct way. However, there are also profound connections. Recall Pickard’s idea that clinicians should take the stance of responsibility without blame in interactions with clients. In the present PhD thesis, Article 1 also draws on a notion of stances, yet in terms of decentering. This concerns the stances of disidentification and nonreactivity. However, the stances of disidentification and nonreactivity do not concern *interpersonal* workings between client and therapist. Instead, these stances concern *intrapersonal* workings through which subjects relate to their own mental processes. I think this is an interesting connection. It stresses that stances are involved in a twofold way in clinical practice, namely as social *interpersonal* workings and as introspective *intrapersonal* workings.
Further, recall Bortolotti’s agency-first approach to belief-like attitudes in the clinical domain. I think this approach is present in Articles 1–3 as well, although it might be in a kind of subtle way. Articles 1–3 do not discuss delusional beliefs in conditions like schizophrenia. Instead, these articles discuss how mindfulness can aid patients in controlling attention, emotions, and moods in disorders like depression and anxiety. As just mentioned, this involves the stances of disidentification and nonreactivity. The relevant thing here is that such stances involve certain belief-like attitudes. That is, these stances involve viewing one’s own pathological states as external to oneself, as non-expressive of one’s identity and as inaccurate representations of one’s own life. As Article 1 argues, if these belief-like attitudes are beliefs about the metaphysical nature and status of mental states, then they are philosophically controversial. The express ideas about the identity and ownership of mental states that do not seem sufficiently justified. However, Article 1 argues that we should not understand these belief-like attitudes as such metaphysical claims. Instead, we should evaluate them in terms of the benefits they bring to subjects lives. This benefit concerns regulation of affections, emotions, and moods. Yet, it also concerns more epistemic processes of the client. Disorders like depression and anxiety most often involve heavy biases and epistemic dysfunctions in thinking, perception, and belief-formation. They disturb several epistemic functions of agents. Since the stances of disidentification and nonreactivity downregulate such biases, these stances seem to bring positive epistemic consequences to the subjects’ broader epistemic functioning. Hence, we might conclude that, even though the stances of disidentification and nonreactivity involve belief-like states that might be problematic from a traditional epistemological perspective on the metaphysics of the mind, these stances’ belief-like states seem rational from an agency-first perspective.

Let me end by stressing that this form of philosophy with a therapeutic focus is actually an old approach in the history of philosophy. For example, Hadot (1995) argues that many branches of ancient Greek philosophy approached philosophical practice as a highly practical discipline concerning how to live one’s life well. Stoicism is a good example of such an approach. I think we can say the same about the sources of modern mindfulness practice, namely the Buddhist tradition. Buddhism is mainly a discipline about how to eradicate suffering from one’s own life and the lives of other people (Siderits, 2023). Traditions like Stoicism and Buddhism are highly relevant today. Not only are they popular topics in self-help books, they are also among the primary building blocks of our most efficient contemporary cognitive psychotherapies, widely used in clinical practice (Diaz & Murguia, 2015). Even those terms like metacognition, attention control, and skilled mental action are modern expressions;
it is clear that traditions like Stoicism and Buddhism have delivered fundamental ideas about these phenomena and how they link to each other. For example, they have discussed how subjects must be aware of their own attention and control it properly, and how they must approach their own mental states as nothing but mental states (instead of considering them accurate readouts of reality by default). The present PhD thesis aims to discuss such ideas through the lens of current philosophical psychology. This is an attempt to bring back old philosophical topics by contemporary means.

References


Rosenkranz, M. A., Dunne, J. D., & Davidson, R. J. (2019). The next generation of mindfulness-based intervention research: what have we learned and where are we headed?. *Current opinion in psychology*, 28, 179-183.


http://www.jstor.org/stable/24672956


**Decentering and attention**

1. **Introduction**

Psychologists describe decentering as the mental operation in which a subject ‘moves out’ of immersion in a mental state (Bernstein et al. 2015, 2019). When pathological states occur, such as depressive thoughts or anxious feelings, patients can execute decentering and prevent the states from determining their further state of mind. This makes decentering a, if not the, key element in contemporary cognitive psychotherapy. The majority of contemporary and evidence-based forms of cognitive psychotherapy train this mental operation, in particular third-wave cognitive-behaviour therapies (Hayes et al., 2011).

Despite its prominence in clinical psychology—in its various guises, such as clinical practice, effect research, and theorising—the philosophical literature does not contain any dedicated analysis of decentering. Philosophers have discussed related issues or mentioned decentering in relation to other topics such as mindfulness, mental action, self-control, psychedelics, etc. (e.g., Repetti, 2019; Latham, 2016; Zawidski, 2018; Letheby, 2021). Yet, these analyses have not targeted the decentering construct as found in clinical psychology. They have not analysed this construct in greater detail. This is unfortunate. The basic idea that agents can intentionally ‘move out’ of their immersion in a mental state is theoretically intriguing. Such ‘moving out’ relates to more fundamental philosophical issues of experience, psychological control, and agency. Besides, the relevance of decentering to mental health also makes the operation of great practical interest.

The present paper aims to fulfil two tasks. First, the paper offers a philosophical explanation of decentering. It bases this explanation upon careful analysis of the relevant psychological literature. The paper argues that decentering is a specific psychological operation involving two essential dynamics. That is, when an agent decenters from a mental state, M, she both introspectively attends to and detaches from M. Second, the paper argues that a prominent theory of attention, known as the selection for action theory, cannot describe this dual nature of decentering adequately, despite the fact that decentering is an attention phenomenon. The selection for action theory claims that the nature of attention is that of selection for action, i.e., it concerns selection of one among many possible inputs for coupling this input with one among many possible actions. Accordingly, the theory claims that the nature of attention control is that of structuring such selection through relevant goals or intentions. However, the selection
for action theory struggles to explain the dual nature of decentering. Some of these problems might not be unique to this specific theory but apply to current mainstream philosophical and scientific theories of attention in general.

The paper is structured as follows. §2 provides the philosophical explanation of decentering. §3 argues for two auxiliary claims about the relation between decentering and attention control. §4 outlines the selection for action theory. The section argues that the current resources of the theory seem insufficient in accounting for decentering or involves problems of coherence with relevant background theories. §5 provides broader considerations on the prospects of accounting for decentering within philosophical and psychological mainstream theories of attention.

2. Decentering
Psychologists describe decentering as the mental operation of ‘moving out’ of immersion in an occurring mental state. This ‘moving out’ involves particular dynamics. Decentering has its roots in mindfulness practice. Related, researchers consistently describe decentering as ‘a detached-observer perspective on one’s ongoing internal experiences’ (Naragon-Gainey et al. 2023, p. 1), or ‘the ability to observe one’s thoughts and feelings from a detached view’ (Zhang et al. 2016, p. 147). Scales used to measure subjects’ ability to decenter align (e.g., Bernstein et al., 2015). Psychologists often describe decentering as involving two basic dynamics. That is, when an agent decenters from a mental state, she both introspectively attends to this state and detaches from it (e.g., Wells, 2009; Hayes et al., 2011; Bieling, et al., 2011; Teasdale, 1999; Bishop et al., 2004; Lange & Grünbaum, 2023). Aligning with such descriptions, I propose that we view decentering as a complex mental operation involving the simultaneous execution of two sub-operations. Philosophically, we can define it as follows:

Undergoing a mental state, M, an agent, A, decenters from M if and only if A simultaneously
(a) introspectively attends to M (sub-operation 1)
(b) detaches from M (sub-operation 2)

An example might be illustrative. Consider Patricia who suffers from depression. Patricia enrolls in contemporary cognitive psychotherapy where she trains decentering. This involves Patricia sitting quietly, closing her eyes, and introspecting any pathological mental state that might occur (i.e., any depression-related state such as thoughts or feelings of meaninglessness).
When a depressive state occurs (like the thought ‘my life is pointless’), Patricia is supervised to both increase her introspective attention to the state and detach from it. By detaching, Patricia downregulates the grip the state has on her processing. She creates a psychological room for manoeuvre so that she can relate to the depressive thought more flexibly, involving that she can think, feel, plan, and behave in ways less determined by it (Wells, 2009; Teasdale, 1999).

Let me elaborate on the two conditions in more detail. Sub-operation 1 involves an agent intentionally deploying introspective attention to the relevant pathological mental state. In line with its origin in mindfulness practice, decentering involves that patients become increasingly aware of the relevant state (Goleman & Davidson, 2017). Authors also formulate this as introspection increasing the conscious representation of the state (Wells, 2005). Decentering is, hereby, profoundly different from other therapeutic strategies. Strategies of distraction or dissociation involve that patients turn their attention away from the pathological mental state and to other stimuli. For example, patients might cope with their depressive thoughts or feelings by re-directing attention to neutral external objects in the environment (as in Attention Training Technique, see Wells & Matthews, 1994).

Sub-operation 2 of decentering involves that an agent intentionally detaches from the pathological mental state. This means that an agent downregulates the influence the state has on her psychological processing. Psychologists take the detachment of sub-operation 2 to involve that agents engage in the dynamics of disidentification and non-reactivity (Bernstein et al., 2019). The process of disidentification involves that an agent relates to the occurring mental state as external to herself and not as an expression of herself (Lau et al., 2006). For example, an agent transfers from thinking ‘my life is pointless’ to relating to the state as outside of herself, as in the observation ‘the thought “my life is pointless” occurs’. Disidentification is the process of establishing a more observational, third-person perspective on occurring mental states (Hanley, Dorjee & Garland, 2020).

---

8 Some authors might find condition (a) mistaken and argue that decentering involves other kinds of self-awareness than introspection (e.g., Lutz et al. 2015; Dunne, Thompson & Schooler 2019). Yet, the above definition best captures the idea of decentering found in third-wave cognitive-behavior therapies and their clinical protocols, which is focus of this paper.

9 Notice that condition (a) and (b) demand that the subject intentionally executes introspective attention and detachment. Some readers might find this controversial. For example, Letheby (2021) describes cases of psychedelic induced decentering. Likely, such decentering does not involve intentional introspection or detachment. Although this is a relevant consideration, it does not pose a problem to the above definition. First, the definition is intended to describe decentering as researched in clinical psychology. Here, authors commonly view decentering as an intentional operation. Second, if one insists that decentering can be unintentional, one could simply distinguish between active (or intentional) and passive (unintentional) forms of decentering. In that case, the above definition only covers active decentering.
Philosophical readers might wonder how one should interpret this psychological idea of disidentification in relation to philosophical discussions about the identity, subjectivity, and ownership of mental states. One interpretation would be to understand the idea in more metaphysical terms. This approach would elaborate disidentification as a transfer in fundamental features of the involved mental states. One could draw on Frankfurt’s (1976) theorising on identification and externality here. Frankfurt distinguishes between active and passive mental states. If a mental state is active, then the agent is engaged in it. An agent can be identified with active states—these states are expressive of the agent. Passive mental states are those that merely happen to an agent. An agent is a passive bystander to these states. A clear example of a passive mental state would be one induced directly by another external agent from the outside (e.g., a neuroscientist inducing a certain thought in you by the use of a neurotechnology). One might propose that when an agent, A, disidentifies with a mental state, M, she transits from being identified with M to M being external to A.

This approach is problematic, however. One major reason is that it might be implausible by Frankfurt’s own theorising. The exact conditions for when an agent can be identified with a mental state, and when the state is external to her, are difficult to specify. Nonetheless, an agent’s attitude towards an occurring mental state is (usually) not sufficient to make the state external to her (Frankfurt, 1976). For example, if an agent disapproves of the reoccurring desire to cheat on an exam, this attitude of disapproval is not sufficient to make the desire external to her. Unless we have a compelling reason for why disidentification involves something radically different, it makes it insufficient for turning a pathological state into an external state.

However, notice that Frankfurt’s theorising on identity and externality occurs on the background of clarifying the agency and responsibility of mental states. The philosophical literature contains alternative approaches to the identity of mental states. Another line of theorising investigates whether subjectivity, or ownership, is a necessary condition for a mental state to be conscious. Numerous philosophers accept this idea (Eilan, 1995). An alternative metaphysical interpretation would be to understand disidentification as opposing this mainstream idea. Disidentification would involve a metaphysical transfer from a patient occupying a mental state, to this state occurring without the patient, or any subject at all, occupying it.

Some philosophers might be sympathetic to this interpretation (Metzinger, 2004). Yet, we are not forced to interpret psychologists as making this, or a similar, metaphysical claim. Psychologists describe disidentification more as a way of relating or approaching one’s own states (Hanley, Dorjee & Garland, 2020). Claims about the metaphysics of consciousness and
claims about how to relate to one’s own states are obviously on different levels. It makes perfect sense to claim that a patient can relate to her occurring mental state as external to herself, without this implying any claim about a metaphysical shift in the actual identity, subjectivity, or ownership of the state.

I refrain from strong metaphysical interpretations in the following sections. Here, I think we should understand the disidentification of sub-operation 2 as a practical stance in relation to one’s own mental states. I think this aligns best with the view of clinical psychologists. We should not consider the ideas of identification and externality as a claims about the metaphysics of ownership of mental states, but as claims about the dynamics of psychological control. This relates back to Frankfurt (1976). Frankfurt (1976) suggests that agents can subjectively identify with an occurring mental state or view it as external, independently of whether the state is in fact metaphysically identifiable or external to the agent. That is, an agent can deliberatively, from her own subjective point of view, relate to a mental state as an expression of herself or as something external. This dynamic is decisive for the agent’s further state of mind. If an agent subjectively identifies with a mental state, she views the state as a meaningful expression of herself. This motivates taking the state seriously, which makes it more prominent in governing her mind and associated actions. By contrast, if an agent subjectively views a state as external to herself, she (probably) does not view the state as meaningful or worth taking seriously. Hence, such subjective externalisation (most likely) decreases a state’s influence on one’s mind and associated actions, compared to subjectively identifying with the state (Frankfurt, 1976). I think we should view disidentification as exactly this kind of regulative dynamic. This analysis aligns with those of other philosophers, as Latham (2016), who also view decentering practices like mindfulness as executions of psychological control.

This analysis also highlights how the disidentification of sub-operation 2 differs from the kinds of depersonalization found in conditions like dissociative disorders and schizophrenia. These conditions involve individuals experiencing occurring mental states with a disturbing vagueness, or as belonging to other individuals than themselves (Sass et al., 2018). The disidentification of sub-operation 2 is not a psychiatric condition involving involuntary experiences of vagueness or alienation. Instead, it is an active stance towards one’s own pathological mental states. It is something that patients can learn, train, and voluntarily deploy, without this involving vagueness or alienation.10

10 Although extreme decentering practices can induce dissociative experiences (Lindahl et al., 2017).
This brings us to the second aspect of sub-operation 2, namely that of non-reactivity. Non-reactivity means that a subject does not act bodily on the basis of the relevant state, e.g., that she does not seek solitude on the basis of a depressive thought or use a substance on the basis of a desire. Moreover, it also means that the subject avoids any mental interference with the state. Interference refers to processes such as evaluation (e.g., negative assessment of the state, as in ‘I should not think/feel like this!’), cognitive elaboration (e.g., rationalising the state as in ‘this thought occurs because event X happened’), or evaluating the state as accurate (e.g., as in implementing a depressive thought as a point of departure for further reflection).

This aspect of non-reactivity makes decentering different from other therapeutic procedures of cognitive psychotherapy. In older forms of cognitive-behaviour therapies, therapists often provoke patients to evaluate the accuracy of their pathological thoughts and emotions through critical reflection (Ellis, 1962; Beck, 1976). The non-reactivity of sub-operation 2 is a way of avoiding such evaluation and reflection (this would be a kind of interference). To use an analogy, the best way to silence destructive, illegitimate political voices is sometimes to remain non-reactive to their assertions. This might be the most appropriate route to hindering that they frame public discourse. The idea is the same for the non-reactivity of decentering.

These aspects of disidentification and non-reactivity enable detachment from a mental state. They downregulate the state’s determination of further processing, meaning that the state is less determinative in a subject’s planning, thinking, feeling, perception, and behaviour (Narogan-Gainey et al., 2023). Let me also stress that disidentification and non-reactivity are closely linked and mutually supportive: establishing disidentification often involves remaining non-reactive and vice versa (Bernstein et al., 2019). In fact, one might think that they directly imply each, or that they can be reduced to one more basic stance towards one’s own mental states. It is beyond the present scope to discuss this in detail. In terms of defining decentering, I shall understand disidentification and non-reactivity as both sufficient for detachment. This means that the second condition, condition (b) of the above-stated definition of decentering, is an inclusive disjunction. That is, an agent detaches from a mental state if and only if she succeeds in the stance of disidentification or non-reactivity towards it.

Some sceptical readers might object to this reasoning. They might stress that reference to clinical practice and patient reports is not sufficient to establish the causal claim that the stances of disidentification and non-reactivity downregulate the relevant mental states. We need more behavioural evidence. Although I think it is correct that such evidence would bolster this causal claim, such an objection unjustifiably neglects the importance of descriptions of
clinical practice and patients reports as sources of evidence. As stressed above, contemporary cognitive therapies consider decentering to be an operation of downregulation. In the clinical literature, subjective reports often stress that even standalone and brief exercises of decentering enable downregulation of negative affective states, emotions, and mood (Huffziger and Kuehner, 2009). Chronic pain patients also report that decentering enables downregulation of pain states’ determination of their mental life and behaviour (Poletti et al., 2021). Further, there is also relevant behavioural evidence. Studies suggest that decentering reduces the determination of craving states on subjects’ cognition and behaviour. Consider that mindfulness meditation decreases automatic food bias as measured in reaction time in categorisation tasks and physiological responses such as saliva volume (Papies, Barsalou and Custers, 2012; see also the review by Keesman et al., 2017).

Alternatively, other readers might stress that authors often describe decentering simply as ‘bare’ attention, i.e., a way of attending to one’s own mental states without identifying, evaluating, interfering, or acting upon these states (see discussions of mindfulness, Puc, 2019). Following, these readers might stress that it is more adequate to define decentering by one condition of bare introspective attention, instead of two conditions of respectively introspection and detachment.11

One could answer this criticism in two ways. First, one could approach the issue merely as a terminological dispute. Common descriptions portray bare attention as attention without identification or reactivity (Gethin, 2011). This sounds exactly like introspective attention of a mental state with the stances of disidentification or detachment towards the state. Hence, it seems to be the case that the stances of disidentification and non-reactivity are packed into the notion of bare attention. It is merely a terminological dispute whether you prefer the one or two condition definition of decentering.

Second, one could reply that there is a more substantial reason for preferring the two condition definition above the one condition version. Overall, the two condition version is more illuminating. The two condition version clearly shows why decentering is something more than merely introspecting one’s mental states. Generally, it seems evident that introspection and detachment can come apart. A subject can introspect her mental state, without detaching from it (e.g., when I introspect whether my desire for coffee is stronger than my desire for tea). Further, a subject can also detach from her mental state without introspecting it (e.g., when I

11 I am grateful to an anonymous reviewer for stressing this issue.
detach from my depressive thought by retrieving a happy memory, see also §3). The two condition version clearly acknowledges this possible dissociation.

Moreover, the two condition version coheres better with standard psychological theories of decentering. These theories also view decentering as a complex operation where multiple, inter-related sub-operations work together (e.g., the Metacognitive process model, Bernstein et al., 2015). I think there is a strong, theoretical reason for breaking down decentering into two sub-operations. The psychological literature views detachment as a form of psychological control (Jankowski and Holas, 2014). By contrast, authors tend to view introspection more as an epistemic operation of merely representing one’s own states (Schwitzgebel, 2010). Collapsing the dimensions of introspection and detachment into one condition of bare attention plausibly prompts the mistaken idea that decentering is exhaustively a matter of an epistemic change in how one introspects. Defining decentering in terms of both introspection and detachment brings out the aspect of psychological control more clearly.

3. Two auxiliary claims

Let me proceed to articulate and defend two auxiliary claims about decentering and attention. These are claims about how to understand decentering in terms of downregulation and attention control. They provide the basis for the following critical discussion of the selection for action theory. The first claim is:

(1) When an agent, A, decenters from a mental state M, A intentionally and directly detaches from M.

I expect clinical psychologists to accept (1), but let me briefly clarify and motivate the claim for sceptical readers. Recall that the detachment of sub-operation 2 refers to the process in which an agent downregulates a mental state’s determination of her processing. This detachment or downregulation is intentional (Bernstein et al., 2015, 2019). This is not the context to develop a fine-grained analysis of what intentional detachment would involve. In very rough terms, intentional detachment would be the process in which an agent’s downregulation of a given mental state is caused, structured, or guided (in the right way) by the agent’s goal or intention (following discussions of intentions in philosophy of action, Grünbaum, 2003; Buehler, 2019).

The intentional detachment of decentering is direct. We can understand this in contrast to indirect detachment. Indirect detachment would involve that an agent detaches from a mental
state, M, by up-regulating another mental state, M*, such that her processing is determined by M* with the indirect effect or by-product that she detaches from M. Contemporary cognitive psychotherapies do utilise such indirect detachment (Wells & Matthews, 1994). For example, having a depressive thought, a patient could intentionally recall a memory of which she is deeply grateful. The patient could also increase her perceptual attention to external objects in the surrounding environment. Both manoeuvres would indirectly detach and downregulate her depressive thought.

Yet, the detachment of decentering is not indirect, it is direct. In direct detachment, the downregulation of a mental state is not a side-effect of the up-regulation of another mental state. The detachment is itself the instant operation performed by the agent. Describing the detachment of decentering as direct is the best way to make sense of the instructions and the phenomenology of decentering. Recall that patients receive the instruction to detach from their pathological mental state through deidentification and non-reactivity. Deidentification and non-reactivity are ways of relating to the pathological mental state. They are not ways of relating to any other mental state. Patients receive no instructions in upregulating other such mental states. Neither does decentering involve the experience of upregulating another mental state. It involves the experience that the pathological state remains present but that there is an increased distance to it (Wells, 2005, 2009). This aligns with psychological theorising. Several researchers have stressed that paradigmatic practices of decentering, such as mindfulness meditation, train the ability of individuals to intentionally prevent stimuli or mental states in determining their processing. These psychologists describe this as a psychological operation of inhibition in the form of direct detachment (Lee & Chao, 2012; see also Nigg, 2000).

Again, some readers might object that we need evidence of a more objective kind to support the hypothesis that sub-operation 2 is a form of direct detachment, and to rule out the alternative hypothesis that it is a form of indirect detachment. Such readers might stress that these are mechanistic hypotheses, they make claims about how processing actually unfolds. Reference to clinical practice, patient reports, or higher-level theorising on mindfulness is not the right kind of evidence. I obviously think that such evidence would be highly desirable, though I do not think that it is required. The claim of (1), and the claims of this paper generally, involves a discussion of how to adequately conceptualise the nature of attention and attention control. This level of theorising is common in the philosophy of attention (Jennings, 2012; Garneri, 2017; Watzl, 2017). Notice that claim (1) is not a claim about the nature of neural inhibition. Some authors argue that all neural inhibition is indirect, i.e., it is side-effect of excitatory mechanisms. Other authors believe that direct neural inhibition does occur, stating
that the mind, or the brain, does involve mechanisms of downregulation without these being side-effects of excitatory mechanisms (see Badre, 2020, ch. 6). Claims about neural inhibitory mechanisms and claims about psychological detachment concern different levels, and it is not obvious that there exists any entailment relation between them. Further discussion of this topic is beyond the present scope, yet it does seem worth future work.\footnote{Relatedly, some readers might think that the idea of ‘direct detachment’ makes controversial philosophical assumptions on the mental agency of attitudes. For example, Hieronymi (2006, 2009) distinguishes between evaluative versus managerial control of attitudes. In short, the first one is control of attitudes by changing one’s answer to the question an attitude embodies, while the second one is a matter of acting upon one’s attitudes by indirect means (e.g., expose oneself to evidence or certain reasoning). An anonymous reviewer suggested that agents plausibly control their stances of disidentification and non-reactivity through managerial control, i.e., indirect means. This might be the case. Yet, it would not be a problem to the above claim of (1). The distinction between indirect and direct detachment concerns the issue of whether an agent downregulates a mental state as a side-effect of upregulating another state (indirect detachment) or without such upregulation (direct detachment). The claim that decentering involves direct detachment entails no commitment to a claim about whether agents establish the stances of disidentification or non-reactivity through managerial control or not. One can claim that these stances are established by managerial control, yet also claim that they directly detach the relevant mental state that they target. Consequently, the claim of (1) does not subscribe to any general voluntarism about attitudes.}

Let us proceed to the second auxiliary claim, namely:

(2) Intentional and direct detachment is sufficient for attention control.

By this, I mean that when an agent executes intentional and direct detachment, this qualifies as the agent executing attention control. Let me develop and motivate (2) by reference to three psychological literatures.

In clinical psychology, researchers view the training of decentering as essentially the training of a patient’s ability to control her attention (Moore & Malinowski, 2009; Wells, 2005, 2009). The detachment of decentering enables patients to avoid attention patterns that underlie their illness, because they can downregulate the mental states that constitute this pattern. Upon such successful downregulation, patients can instead allocate their attention in ways fruitful for their well-being—such as directing it for pursuits of valuable projects, generation of productive emotions, engagement in social relations, etc. This is how the clinical literature conceptualises intentional, direct detachment as a kind of attention control.\footnote{The question “what is attention?” is obviously controversial in both the philosophical and scientific literature. Let me stress that clinical psychologists do not make any commitments to specific theories on this matter—neither does the above argumentation presuppose such commitment. Instead, clinical psychologists treat attention in accordance with Mole’s (2021) basic, pre-theoretic characterization that attention concerns the selective directedness of our minds. The issue of competing attention theories concerns the exact nature of this directedness.}
Decentering is also a research topic in sports and performance psychology. Psychologists study decentering in relation to an athlete’s ability to control her attention before and under performance (Brick, MacIntyre & Campbell, 2014). This relates to a common distinction. Sports and performance psychologists generally distinguish between external and internal distraction. External distractions refer to immediate environmental objects that disrupt the attention of athletes (e.g., salient visual or auditory stimuli). Internal distractions refer to an athlete’s own mental states that disrupt her attention (e.g., her own on-going worry, rumination, strong affective states, etc.). Internal distractions often pose the greater threat against an athlete’s attention. Here is psychologist Aidan Moran’s description of attention lapses in intense performance situations:

A classic example of a costly internal distraction occurred in the case of golfer Doug Sanders, who missed a putt of less than 3 feet that prevented him from winning his first major tournament… Remarkably, Sanders’ attentional lapse was precipitated mainly by thinking (Moran, 2012, p. 122).

Studies have investigated whether the training of decentering (through mindfulness meditation) can aid athletes in controlling internal distractions. Result suggest that this is often the case (Cathcart, McGregor & Groundwater, 2014; Moran, 2012). Consequently, many sports and performance psychologists would agree with claim (2). They view decentering as the intentional act of minimizing internal distractors, i.e., a way for athletes to control their attention.

Lastly, you can make an argument for (2) by drawing on considerations from cognitive psychology. It is standard to view executive control as having a three-fold nature. These three aspects are updating of working memory, task shifting, and inhibition of irrelevant information (Miyake et al., 2000). The link between the notions of executive control and attention control is intimate (Buehler, 2019; Watzl, 2017). One might even think that these notions refer to the same basic capacity (the fundamental ability of agents to arrange and maintain their own processing in accordance with their relevant goals). If you accept that agents can perform intentional and direct detachment, then this surely seems to be an act of executive and attention control (Gallant, 2016).
4. Selection for action theory

Consider now whether the selection for action theory of attention (SfA) has the conceptual resources to account for decentering satisfyingly. Let me first provide some background. Overall, the topic of attention concerns the selective directedness of our minds (see also endnote 6). Yet, theories of attention disagree about the nature of this selectivity. One dominant theoretical approach views attention in terms of the capacity limitation of cognitive systems (Broadbent, 1958). The central idea is that since cognitive systems, as the human brain, are limited in their ability to process information, they must contain a mechanism that can filter certain information above other. Attention is this mechanism. In metaphoric terms, attention functions like a bottleneck.

The selection for action theory (SfA) does not view attention in terms of the limited capacity of cognitive systems. Instead, SfA views attention as a selection mechanism handling the excess of information in cognitive systems (Mole, 2021). Neumann (1987) and Allport (1987) originally proposed this idea in cognitive psychology. They theorised that since cognitive systems can process multiple information simultaneously, the systems must select among the information if they are to ensure coherence and coordination in their action. Since then, SfA has been widely successful in psychology and cognitive science. Some authors claim that it is the standard, scientific conceptualisation of attention (Wu 2011a, 2022; although see Buehler, in press).

Wu has provided a philosophical elaboration of SfA (2011a, 2016, 2022). The notion of the many-many problem is central in this formulation. The many-many problem refers to the fundamental situation of agency that agents always receive multiple inputs (e.g., perceptual inputs or thoughts), which they can couple with multiple outputs (bodily or mental responses). To do anything, agents must establish a behaviour path by selecting an input for coupling it with an output. In bodily action, the output is a bodily activity; in mental action, the output is a non-bodily activity (Wu, 2022, p. 70). The central idea is that attention is a solution to the many-many problem: an agent’s attention is her selection of an input for an action. Moreover, according to SfA, attention control is the way in which an agent’s goals or intentions top-down structure this selection of an input for action (Wu 2011a, 2016, 2022). Hence, we can outline the two central theses of SfA as follows:
**SfA—attention.** An agent, A, attends to an input, P, if and only if A selects P for action.

**SfA—attention control.** An agent, A, executes attention control if and only if A’s relevant goals or intentions structure a selection of an input for action.

It is crucial to stress that **SfA—attention control** claims that the job profile of attention control is fully *restricted* to this particular role of structuring selection of an input for action. Metaphorically, attention control serves the role of traffic management, like that involved in train traffic (Wu, 2016, p. 106-7). Attention control is the management of which input among many (i.e., which incoming train) is coupled with which output among many (i.e., which track the train is to continue on). Like the train manager’s job is exhaustively a matter of selecting the incoming trains for an outgoing track, without directly steering the train, attention control is restricted to the structuring of the selection of an input for action, not the direct execution of the action (Wu, 2016, p. 109).

### 4.1 Challenge

We can pose the challenge to SfA now. Consider decentering from a general perspective. Decentering involves attention to a relevant pathological state. This type of attention does not seem to be a matter of selecting the pathological state for an action. It seems to involve the opposite. Decentering is a matter of attending to a state with the purpose of it *not* being selected for an action. Hence, we might say that decentering involves attending to a mental state with the purpose of distancing from it, or with the purpose of non-action. Furthermore, decentering poses a more specific challenge. Accept for now the two auxiliary claims about decentering: (1) the detachment of decentering is intentional and direct, and (2) the detachment is an execution of attention control. It is difficult to see how the thesis of **SfA—attention control** could have the conceptual resources to account for the intentional and direct detachment as a form of attention control. According to **SfA—attention control**, attention control is always a case in which an agent’s goals or intentions structure the selection of an input for action. However, in line with the remarks just made, the intentional and direct detachment of decentering involves targeting a mental state for downregulating it. As an act of attention control, decentering is the operation of avoiding that a state produces or influences action.

The challenge is then that decentering seems to involve dynamics of attention and attention control that cannot be described adequately in terms of selection for action. Other philosophers have argued against SfA’s necessary and sufficient conditions for attention and
attention control (see Buehler, in press, for an overview). The present challenge concerns necessity. It argues states that some mental episodes do not fulfil the requirements of SfA—attention and SfA—attention control, yet they are instances of attention and attention control.

SfA supporters might respond in various ways. I think the most promising strategy is for the SfA supporter to acknowledge that from a holistic perspective, decentering does appear difficult to account for under SfA. However, when we split decentering up into its two sub-operations, the supporter might argue that SfA can indeed account for the involved aspects of attention and attention control. Let us explore how the SfA supporter could elaborate and defend this reply.

Recall that sub-operation 1 concerns introspective attention to the pathological state. The SfA-supporter might propose that introspective attention is the selection of an input for the action of increasing its conscious representation or salience (as Wu proposes in cases of perceptual representation without motor activity, see 2011b, p. 105, 2022, p. 72). In terms of sub-operation 1, this involves the attention of selecting a pathological state for increased conscious representation, and it involves the attention control of structuring this selection. Notice that some philosophers have argued that it is hard to make sense of increase in conscious representation or salience as a form of action (e.g., Watzl, 2017). I shall not assume anything about this general dispute here. Our main focus will be on sub-operation 2.

Recall from §3 that sub-operation 2 involves the intentional and direct detachment of a mental state, and that this is an execution of attention control. The pressing task for the SfA supporter is to show how we can understand this detachment through the thesis of SfA—attention control. This is the task of showing that we can adequately describe sub-operation 2 as an agent’s intentions or goals structuring the selection of the pathological state for some action. I think the SfA supporter should elaborate this idea in one of the following three ways—although I also think that none of these succeed in the end.

4.2 Selection for meta-representation
We could understand sub-operation 2 as the operation in which agents intentionally meta-represent their pathological states. That is, disidentification and non-reactivity are ways to represent or ‘tag’ a depressive thought as ‘external to me/not an expression of me’ or ‘not to be reacted upon’. This involves the structured selection of the pathological thought for the action of meta-representing it.

However, such meta-representation would plausibly not be sufficient for the detachment involved in sub-operation 2. One can imagine an agent genuinely meta-representing her mental
state as ‘external to me/not an expression of me’ or ‘not to be reacted upon’ without this involving that she is actually able to detach and downregulate it. A parallel could be that one can represent an external object as ‘not to be acted upon’, without being able to control one’s action around it. For example, one might genuinely represent a cake as ‘not to be eaten’ or a looming object as ‘not to be ducked’ and yet not be able to act accordingly. My point is that the detachment, facilitated by disidentification and non-reactivity, seems to involve some kind of psychological control that is not exhausted merely by meta-representation. This accords with the clinical literature. Although some psychologists think of decentering practices as involving meta-representation, they also stress an additional and crucial aspect of psychological control (Jankowski & Holas, 2014).

The SfA-supporter might object that this reasoning neglects the fact that many philosophers and psychologists believe that representations can be of non-propositional format, such as sensory or motor format (Pacherie, 2008). Building upon these general views, the SfA-supporter could argue that the meta-representation involved in sub-operation 2 is of non-propositional format such that if a mental state is encoded in this format, it ensures detachment (perhaps similar to an idea of motor representations which claims that information encoded in such representations has direct consequences on motoric activity). Before we can properly assess such suggestion, the SfA-supporter would have to elaborate on this idea of non-propositional meta-representation. It does not seem to follow from common views in the clinical literature and related psychology on meta-representations as involved in operations like decentering (Jankowski and Holas, 2014; although see also Dunne, Thompson, and Schooler, 2019). Moreover, even if one could formulate a plausible account of such non-propositional meta-representation, §4.4 presents a general objection (concerning the maintenance of stances) that would apply to this proposal as well (more about that below).

4.3 Selection for turning off
Another reply would be that sub-operation 2 involves selection of the pathological state for the action of turning it off. One way to fit the stances of disidentification and non-reactivity into this picture would be to view them as standing states or attitudes that structure this selection of a state for turning it off.

Yet, there are several problems with this idea. One major problem is that the account does not fit well with the computational background theory of SfA. For example, Wu (2016, p. 107) develops SfA—attention and SfA—attention control by drawing on the biased competition view of neural networks. This view considers neural networks to be landscapes of
multiple input-output mappings. The mappings compete in being realised. The higher the weight of a mapping is, the more probable it is that the mapping is realised. The causal role of intentions is to bias weights. This means that an agent controls her attention (i.e., the relevant input-output coupling) when this coupling is appropriately biased by the agent’s relevant intention. Importantly, according to one prominent understanding, this biased competition only involves *indirect* inhibition. Inhibition of an input-output mapping, M, is always an indirect effect of M losing in the competition to another mapping, M*. Inhibition is never a direct turning down of the weights of one mapping. It is the turning up of the weights of some other competing mappings (Hommel, 2000).

With this background theory in mind, the present proposal seems to involve two interrelated problems. First, selection of an input (i.e., a pathological state) for the action of turning it off sounds like the kind of direct inhibition that the biased competition view rejects. Second, by this proposal, the selection of a pathological state for turning it off would qualify as a selection for action process. Yet, Wu (2011, 2016, 2022) normally describes selection for action processes differently, namely: an input is selected for action, only if it implies that the input produces further effects in processing (see also Wu, 2022). Selection of a state for turning it off involves the opposite. It involves that a state is selected for not producing further effects.

Of course, the SfA-supporter might accept the idea of direct inhibition and also agree to view selection for ‘turning off’ as a selection for action process. However, these modifications would qualify as substantial elaborations of SfA and might not come without problems. They could disturb the coherence between SfA and the background theory of the biased competition view. They also appear to stretch the criterion for what counts as a selection for action process significantly. Nonetheless, we shall not dive further into these issues here. Instead, as mentioned above, I shall give priority to an objection about the maintenance of stances. This objection also applies to the present proposal concerning selection for turning off, irrespectively of whether the SfA-supporter accept the above elaborations.

4.4 Selection for stance

The SfA supporter might propose that sub-operation 2 is the selection of a pathological mental state for the detaching stances of disidentification or non-reactivity. Simply, sub-operation 2 is the selection of a state for a stance. The supporter might stress that a stance is not to be understood as merely a meta-representation or as the output of turning a state off. A stance is more like an active attitude. We might draw an analogy. When accounting for bodily skill, SfA states that the role of attention control is that of coupling an input (say a cup) with a certain
motor output (lifting the cup) (Wu, 2011a, 2016). Even though disidentification and non-reactivity are not skilful motor outputs (as lifting a cup), they can be seen as skilful mental actions. They are ways of manipulating one’s own mental states, just as the lifting is a way of manipulating the external object of the cup.

This reply might seem promising. However, it is problematic for the general reason flagged above. This relates to previous criticisms of SfA. Other philosophers have already argued that SfA’s terminology—of ‘selection’, ‘inputs’, ‘outputs’, ‘actions’, etc.—can be interpreted and applied in such an elastic way that it is seemingly flexible enough to encompass any kind of mental phenomenon or dynamic. This is problematic. It decreases the precision of SfA and its ability to illuminate the nature of attention in a comprehensible way (Buehler, 2018; Watzl, 2017). Let me show how this general criticism pertains to the present proposal of selection for a stance.

Here is the problem. When executing sub-operation 2, subjects undergo periods of time where they maintain their stance towards the pathological mental state. By maintenance, I mean that subjects hold their stances of disidentification or non-reactivity towards the pathological state uninterrupted over a period of time. This might demand some training but it is certainly possible (see descriptions in Wells, 2009; Moore & Malinowski, 2009; Poletti et al., 2021).

It is not obvious that we can adequately describe these periods of maintenance of stances in terms of selection of the pathological state for these stances. Normally, we use the term ‘selection’ to refer to a scenario in which an object is prioritised (often for some further purpose) at the cost of other available objects. This is also the basic meaning of the term in SfA. Then, recall Patricia who suffers from depression. Imagine that Patricia becomes skilled in decentering through extended practice. She can now maintain a stance of non-reactivity towards her depressive thought for extended periods, say 20 seconds, before she becomes distracted or immerse into rumination again. For the sake of the argument, I am ready to grant that on both sides of the 20 second maintenance of stance, we could describe Patricia’s mental life as that of selection of the pathological state for the stance of non-reactivity. Before the maintenance, Patricia must prioritize the pathological state at the cost of other available inputs (as in other thoughts, feelings, or external stimuli). After her stance is disrupted and distracted, (re)selection the pathological state for the stance is necessary again.

However, the phase of maintenance of the stance itself involves something beyond simply selection of a state for this stance. When Patricia is maintaining her stance towards a mental state, her mental operation is not primarily a matter of selection of the state for this stance. It is primarily a matter of her executing, or we might say acting, the stance. Patricia’s
actively establishes and upholds the stance towards the state. In the terminology of SfA, Patricia directly performs the mental action, which the pathological state is selected for—she realises this output. It is a mental action of directly relating to the pathological state. This is also how subjects, trained in decentering, describe this stage of maintenance. These phases concern actively executing the stances. This is what they do. Further, it is by actively executing the stances that the subjects manipulate the grip the pathological state has in their processing and downregulates it. This is how sub-operation 2 involves phases that are not adequately described as processes of selection of pathological states for stances, but direct executions of the stances (e.g., Poletti et al., 2021; Goleman & Davidson, 2017).

However, the SfA supporter might insist that we can describe maintenance of stances as selection for action, drawing on the following analogy. Consider the rules of freeze dance. Every time the music stops, you have to take a stance and hold it for some time (e.g., 20 seconds). SfA can explain this maintenance of bodily stance as a matter of an input being selected for action. Holding one’s bodily stance is a matter of selecting the proprioceptive input associated with the stance one had when the music stopped, and to couple it with the action of keeping the bodily stance associated with this proprioceptive input. This is an active manoeuvre: it demands that no other input is selected (other perceptual inputs, thoughts, etc.), and no other output is realised (e.g., other impulses to move). The SfA supporter might propose to view sub-operation 2 as a mental freeze dance. It concerns freezing one’s mental life by hindering pathological states spreading, like freeze dance is about hindering motoric impulses spreading. Maintaining one’s stance, in sub-operation 2, is nothing more than continuously selecting a pathological state for the action of freezing.14

This analogy has many problems. Overall, the analogy mistakes the goal of decentering as that of being static, or as in freezing one’s mental life. This is incorrect. Decentering is not about freezing or keeping your pathological state constant. It is about continuously downregulating it. Upon successful decentering, the pathological state disappears over time and there is no need to introspect or detach from it.

This also means that the analogy reduces the stances of disidentification and non-reactivity to a form of static maintenance of the pathological state. I acknowledge that the term ‘stance’ might prompt the idea that disidentification and non-reactivity aim for a static state,

---

14 I am grateful to Thor Grünbaum for proposing this analogy.
yet this is misleading. The notion I draw on aligns with that of van Fraassen’s, when he proposes empiricism to be a form of stance and writes\textsuperscript{15}:

> A stance consists of a cluster of attitudes, including propositional attitudes (which may include some factual beliefs) as well as others, and especially certain intentions, commitments, and values. (van Fraassen, 2004, p. 128)

Here, stances are not reducible to a set of factual claims. They are more like complex approaches to epistemic issues. Chang (2022) draws on this idea in his formulation of what he calls realism for realist people. Chang argues that the realism embedded in scientific practice is not a matter of committing to the fundamental claim that our scientific theories are true. Instead, the realism of scientific practice is a way of approaching epistemic issues, i.e., it is a stance. As a stance, it involves intentions, commitments, and values. It also contains forms of know-how and practices that are socially and historically transmitted (e.g., the ability to set up experiments, approach data, etc., as taught through scientific education). Consequently, maintaining one’s stance of empiricism or realism throughout epistemic discourse is not exhaustively a matter of committing to certain factual beliefs. It is a matter of maintaining this complex stance as one’s perspective on the relevant issues.

It is the same notion of stance I draw on. Disidentification and non-reactivity are clusters of attitudes. They involve both passive and active control aspects. Subjects must have the ability of avoiding committing to pathological states (a more passive ability), yet ensure that they uphold this avoidance despite it being very difficult (a more active ability of not immersing into rumination, worry, etc.). The stances also contain central epistemic aspects. When patients train disidentification and non-reactivity they are provoked to adopt and comprehend certain epistemic attitudes towards their own pathological mental states. For example, patients must transfer from viewing these states as accurate read-outs of reality to viewing them as fallible and idiosyncratic. These aspects are taught and practiced in social environments. Patients apprehend them in clinical settings with therapists (and perhaps other subjects), and do so under the general purpose of supporting mental health and well-being (Wells, 2009; Poletti et al., 2021; Goleman & Davidson, 2017).

\textsuperscript{15}Relatively, consider how Pickard (2017) uses the notion of stance in her idea of “responsibility without blame” in clinical work.
Hence, in terms of sub-operation 2, maintaining one’s stance is a matter of ensuring that it is this cluster of attitudes that is perspective by which you relate to your own pathological states. This relation is active, something that an agent does. It is a form of praxis that she has learned and apprehended in various contexts. Nevertheless, there is no recognition of this aspect in the SfA supporter’s analogy. It misunderstands stances as static mental life, when it is in fact an active way of approaching one’s own mental states. The problem is that the minimalist functional terminology of SfA—attention control cannot capture these aspects of executing and maintaining stances as acts of attention control. We cannot describe this execution and maintenance as merely that of structuring selection of a pathological state for some action. These stances are themselves (mental) actions towards the pathological state—mental actions of rich psychological character. Crucially, my claim is not that the maintenance of stances cannot be described as involving input-output couplings or structuring of these couplings, in any way. This would be a much more basic claim against functionalism about this particular operation of the mind.

We can sum up the objection as follows. If the SfA-supporter claims that sub-operation 2 is exhaustively a matter of selection of pathological states for the stances of disidentification or non-reactivity, then she runs into the problem of accounting for the phases of maintaining stances. Yet, if the SfA-supporter insists that these phases can be seen as acts of freezing one’s mental life, then she neglects the purpose, activeness, and richness of maintaining the stances. Notice that this objection applies to the above proposals on selection for meta-representation (§4.2) and selection for turning off (§4.3) as well. Maintenance of stances is a mental action, it is something that agents do towards their own states. Agents execute these stances. This is something more than merely structuring a selection of a state for a meta-representation, or structuring the selection of the state for turning it off. It is a mental action of continuously relating to the state in a certain way.  

The SfA-supporter might dismiss this as a problem. She might claim that although maintenance of stances does involve aspects of attention, it is not an operation of attention control. However, this reply would be problematic. Third-wave cognitive-behaviour therapies understand psychological illness, like depression, as malfunctioning attention patterns (Wells

---

16 Notice I leave one alternative proposal out. This is the reply that sub-operation 2 is composed of two instances of selection for action. Part 1 concerns selecting the pathological state among other inputs for the action of a stance; part 2 concerns the executing of the stance which is itself a selection for action process. However, this account stretches the term of “selection”. If the pathological state is already selected among the other available inputs (in part 1), then what are the competing inputs involved in the selection of the execution of these stances (in part 2)? Moreover, once again, it is not obvious how the account should adequately describe the maintenance of stances.
The protocols of these therapies involve training in attention control (Wells, 2009). It is the execution and maintenance of the stances, which is the central act of attention control at play in sub-operation 2. Enabling patients to maintain stances like disidentification and non-reactivity is an enhancement of their capacities and executions of attention control (Moore & Malinowski, 2009; Wells, 2009). It is by these stances that patients can do the attention control of intentional and direct detachment outlined in §3. It is the mental action of these stances that constitute the central act of attention control.

Decentering and sub-operation 2, hereby, poses a unique challenge to SfA. This challenge is different from other cases. Consider again the case of lifting a cup. Here, we might agree that my attention control is exhaustively a matter of selection of the input of the cup with the action of lifting it. However, in terms of sub-operation 2, the job profile of attention control exceeds the mere selection of a state for a stance, since the maintenance or direct execution of the stance are acts of attention control.

It is important to stress that this is not the same problem as accounting for some mental actions as acts of attention, i.e., accounting for attention as a mental act. Wu (2011b, p. 105) claims that attention can itself be a mental action, namely the action of increasing conscious representation of an input (recall from §4.1 that this is how SfA could attempt an account of sub-operation 1). However, this solution does not transfer to cases as that of sub-operation 2, where the mental action is not that of attention but that of attention control. In the end, we can formulate the novel challenge as follows: maintenance of stances is a mental action of attention control, yet it is not adequately described by SfA—attention control.

Now, the SfA supporter could make two replies. First, she could point to Wu’s (2022) recent terminological modification of SfA. Wu (2022) claims that the slogan ‘selection for action’ is not the best way to communicate the central idea in SfA (at least his version of it). Instead, we should view attention as the process of an input guiding an action (Wu, 2022, p. 66). By this change in terminology, we should understand SfA—attention control as stating that attention control is the structuring of what input comes to guide action. The SfA supporter might argue that this reformulation, somehow, allows to capture maintenance of stances as executions of attention control.

I think this reply is problematic for several reasons. One general problem is that the notion of ‘guidance’ is itself philosophically controversial and difficult to specify, making the reformulation unclear in this respect (until it is coupled with a fleshed-out notion of guidance). Another problem concerns the fact that authors most commonly support SfA by pointing to cases of perception-motor activity. It might make sense to say that an object (a cup) guides a
motor response (the activity of lifting). However, in the case of decentering, it is inaccurate to say that the relevant pathological state guides an agent’s response. Recall that sub-operation 2 involves detaching from this state. Decentering is exactly a matter of hinderling that the pathological state produces, or guides, responses. Finally, even if we assume that the SfA supporter could provide a working notion of ‘guidance’ that could capture this puzzling consideration, it still seems inaccurate to say that maintenance of stances is simply a matter of structuring such guidance. When maintaining their stances, agents do not only structure what input that guides their mental actions. As argued above, agents directly execute a mental action towards the relevant state.

Second, the SfA supporter could draw on Wu’s general ideas on supervenience between processing levels involved in attention (2011a, 2011b, 2016). This is the following idea. Take the example of attending deeply to a piece of music for ten minutes. This temporally extended state of attention might not occur as a continuous selection of the music for some action. Yet, the SfA supporter might argue, we can indeed understand it as such if we view the extended state of attention to the music as a higher-level state that supervenes upon lower-level processes of continuous selection for action. Although one might not experience one’s attending to the music as a continuous selection of the music for some action, this state of attention is indeed contingent upon a selection of the music, occurring continuously at a lower-level.

The SfA supporter might transfer this idea to the present issue on maintenance of stances. Here, the issue would not be about supervenience between levels of attention, but concern supervenience between states of attention control. The SfA supporter might propose to view the maintenance of stances as a higher-level state of attention control that supervenes upon lower-level processes of attention control. That is, while maintenance of stances might not be adequately described as selection for action at the personal-level, this maintenance depends upon continuous selection occurring at lower levels of a subject’s processing.

I think this reply is problematic for two interconnected reasons. First, SfA claims to be a theory that adequately captures attention as a personal-level phenomenon (Wu, 2011a, 2016). Drawing on the idea of supervenience will not support this theoretical ambition. The supervenience reply fails to provide an accurate description at the personal-level, i.e., at the mental life of the subject maintaining her stance. It simply retreats to explaining it in terms of processing at lower levels (see also Watzl, 2017).

Second, the SfA’s reply assumes a very undemanding criterion for when we can label a personal-level mental operation a selection for action process. The reply implies that we are justified in theorising a personal-level mental operation, O, to be a matter of selection for action
if we can make sense of O as supervening upon a lower-level process that we can describe in terms of such selection. This is an undemanding and very thin notion. It is not clear what insight it provides.

I think the above issues connect to the critical point on the elasticity of the terminology of SfA, stressed in the beginning of this section. SfA promises us theoretical progress and elegance when it proposes that all instances of attention and attention control are matters of selection for action. However, to account for sub-operation 2 as a selection for stance, the SfA supporter must stretch this terminology in various ways. Moreover, even if we allowed these terminological manoeuvres, they do not appear to provide insights. I am ready to acknowledge that the present critical considerations might not offer direct conclusive reasons for the claim that SfA cannot account for sub-operation 2. However, they do offer considerable support for the conclusion that SfA, and its thesis SfA—attention control, face serious obstacles in this attempt.

Of course, some SfA-supporters could insist that the stances of disidentification and non-reactivity contain so many ‘non-attentional’ aspects (such as epistemic and control aspects) that it is not reasonable to expect a theory of attention to describe them as acts of attention control. Nonetheless, this reply only highlights the problem. SfA claims that attention and attention control is involved in any intentional action (Wu, 2011, 2016, 2022). Maintenance of the stances of disidentification and non-reactivity are intentional (mental) actions. The claim of this paper is exactly that SfA—attention control cannot adequately conceptualise the attention control involved in these actions. This turns the issue around: since SfA cannot adequately describe the maintenance of stances as acts of attention control, the theory illegitimately excludes certain acts of attention control.

5. Concluding remarks: choices and challenges
This paper proposed to understand decentering as a complex mental operation involving the synchronic execution of introspection and detachment. SfA could analytically approach decentering by targeting its two sub-operations in isolation. In terms of sub-operation 1, the SfA supporter could argue that this involves selection of a pathological mental state for the action of increased conscious representation. Other authors have already provided independent arguments for this being an unattractive account, though this has not been the main concern here. More importantly, the SfA supporter would have to show that SfA—attention control can account for sub-operation 2. This implies showing that we can adequately describe sub-operation 2 as a structured selection of a pathological state for a type of action. I examined
three ways in which the SfA supporter could defend this idea. I argued that all of these involved various problems concerning the dynamics of detachment, coherence with the computational background theory of SfA, and the maintenance of stances.

This paper is not an attempt to provide a conclusive argument concerning the ability of SfA to meet the challenge of decentering. My aim has simply been to introduce a new challenge to theories of attention and explain its philosophical interest. Nonetheless, let me stress that irrespectively of how the SfA supporter attempts an account of decentering, she must accept that decentering, as complex operation, involves agents intentionally manipulating two causal relations of the same mental state at the same point in time. That is, agents both manipulate the causal relation of increasing conscious representation of the state and the causal relation of the state’s determination of further processing. This stresses the puzzling nature of decentering as a multi-dynamical unit of attention and attention control in which agents relate to the same mental state in two different ways. Such multi-dynamical units are not anticipated by SfA.

Finally, let me broaden the perspective a bit. Although this paper has focused on SfA, many of the points seem to have wider relevance. Decentering seems puzzling not only to SfA but to mainstream theories of attention in general. It is clear that decentering challenges a basic assumption widely accepted in the philosophical and scientific literatures. This is the assumption that attending to a phenomenon implies selecting the phenomenon for further processing. That is, when an agent attends to a phenomenon, she upregulates the phenomenon’s influence on her processing. This means that theories generally view attention control as essentially the control of selecting phenomena for upregulating their influence on one’s own processing.

This assumption is, more or less explicitly, endorsed by all mainstream philosophical and scientific theories of attention (Mole, 2021). Consider, for example, the bottleneck theory outlined in § 4. Recall that this theory views attention to be a form of bottleneck. This means that attending to a phenomenon is matter of upregulating the phenomenon’s influence on processing by letting it pass the bottleneck and occupy some of the limited resources of the cognitive system. Attention control is essentially the ability of controlling the bottleneck (Broadbent, 1958). This stresses how the bottleneck theory commits to the relevant assumption and how decentering presents a puzzle to the theory by involving attending to a mental state for downregulating it. It also seems clear that the nature of decentering, as this multi-dynamical unit of attention and attention control, is novel to mainstream theories. Consider the bottleneck theory again. As a multi-dynamical unit, this theory would view decentering as the operation of letting a mental state pass the bottleneck (concerning conscious representation) while
excluding it from passing another bottle neck (concerning further processing). Philosophical or scientific theories of attention do not anticipate this multi-dynamical character of decentering (Mole, 2021). These aspects make it worth selecting decentering for further philosophical processing, independently of SfA.

References


Article 2

Transparency and the Mindfulness opacity hypothesis

1. Introduction

The Transparency Thesis (TT) is widely endorsed by philosophers. Following the philosophical mainstream, we can state TT as a phenomenological thesis:

\[ TT. \text{ When an individual is introspecting their experience, only objects and properties of the objects of their experience can appear to them.} \]

TT claims that individuals cannot be directly introspectively aware of their experiences or properties of their experiences (Harman, 1990). If one tries to gain introspective awareness of one’s experience, one always ‘looks right through’ one’s experience to the object of one’s experience. Phenomenal properties always appear as properties of the objects represented or as objects represented by one’s experience. This means that TT disallows that subjects can become introspectively aware of phenomenal properties that appear as non-object properties, where non-object properties are phenomenological properties that do not appear as representational objects or as properties of these. Although TT is a phenomenological thesis, the thesis is often used as a crucial piece of evidence for prominent metaphysical theories of consciousness and perception, such as strong Representationalism (Dretske, 1997; Tye, 1995, 2000, p. 46-49; Byrne, 2005, 2011; Harman, 1990).

In this paper, we argue that certain forms of mindfulness practise enable a kind of introspective awareness that is in conflict with TT. Other philosophers have suggested that particular forms of attentive awareness trained in some forms of mindfulness practice are best understood as forms of introspection that violate TT (Metzinger, 2013; see also Davis, 2018). We call this the Mindfulness opacity hypothesis (MOH). Yet, MOH has remained underdeveloped in two important respects. First, it remains unclear exactly how introspective awareness under mindfulness is supposed to be in conflict with TT. TT allows that individuals can attend to their experiences and describe them on the basis of their properties. The thesis even allows that subjects can think of phenomenological properties as non-object properties when they are engaged in introspection. It is not clear how MOH opposes TT on these matters.

---

\[ 17 \text{ See for example Kind (2003), Nicholson (2014), and Mitchell (2020).} \]
Second, it remains unclear what would motivate the claim that mindfulness practice can involve a form of introspection that is in conflict with TT. What kind of evidence could support such a claim? We address these two questions.

The paper is structured as follows. In §2, we provide the necessary background for understanding TT. §3 motivates MOH by drawing on statements by clinical psychologists and cognitive scientists about the special features of mindfulness introspection. These statements are evidence for our claim that mindfulness researchers generally endorse MOH. In §4, we examine what MOH implies about transparency and opacity. We argue that MOH could involve shift in experiential perspective, varying degrees of scope, and various levels of skill. In §5 and 6, we examine whether there are any objections open for a TT-supporter that could hinder MOH from getting off the ground.

2. Transparency Thesis
To understand the argument of this paper, it is important to be clear about basic aspects of TT. The central claim of TT is often explained by the analogy with a painting (Harman, 1990). Imagine perceiving Delacroix’ famous Liberty Leading the People. You can perceptually experience or be aware of the properties of the painting in two ways. You might experience the colours and forms as properties of the represented objects in the represented scene. You might perceive the red colour as a part of the French flag and yellow folding structure as the texture of the dress of the woman holding the flag. However, you might also perceive the colours and forms as non-object properties of the painting itself. You might look at the painting in such a way that the red colour and folding lines now genuinely and perceptually appear as properties of the painting. TT claims that introspecting one’s experience is different from perceiving a painting in the following sense. When we introspect our experience, no properties can genuinely appear to us as non-object properties.

The painting analogy necessitates two clarificatory remarks. First, TT is a broad thesis about all kinds of phenomenal states and experiences. Philosophers have often introduced TT as a thesis about introspection of perceptual experience (Tye, 2002; Harman, 1990). TT-supporters have argued that the introspection of such experiences, like that of vision, does not reveal any ‘mental paint’ or qualia. But a more general claim is typically implied. If TT is to count as evidence for metaphysical positions like strong Representationalism, it must cover all kinds of phenomenological characters, including pain experiences and more cognitive states like conscious thinking. We shall approach TT as this broader thesis (Aydede, 2019; Kind, 2003).
Second, the painting metaphor can also help us understand what kinds of introspective awareness TT allows and disallows. Philosophers endorsing TT often consider introspection to be of a rather simple nature. To use the example of pain in the knee, introspecting one’s experience is in some sense simply to prefix the pain experience with ‘I am aware that ___’ (Tye, 2002, p. 138; for a related view of introspection, see Evans, 1982, p. 227). TT allows that individuals can in some sense be aware of their own experiences. It permits that individuals can acquire an increased awareness of the fact that they have this experience of pain in their knee. This would allow that individuals can think about the phenomenal properties of their experience as non-object properties. For example, an individual undergoing pain might think to herself that the involved phenomenal properties (e.g. the stinginess, bodily location, etc.) are not properties of her bodily damage, but instead non-object properties. Of course, TT-supporters like strong Representationalists would claim this kind of thinking to be erroneous, since there are no non-object properties metaphysically speaking.

In contrast to the subject’s cognitive classification of her own experiences, TT is incompatible with a kind of introspective awareness where some phenomenological properties genuinely appear to the subject as non-object properties. We might draw on the painting analogy again to flesh out this difference. You can think about the paint of a painting as a non-object property without this genuinely appearing to you as such (the painting might be a perfectly deceiving trompe l’oeil). When the paint genuinely appears to you as non-object property, the paint property has an experiential nature: you experience the paint and its properties. TT denies that subjects could have an experience in which phenomenological properties genuinely appear as non-object properties in an experiential sense.

The dispute over TT is then the question of whether phenomenological properties can genuinely appear to us in our introspective awareness as properties of the relevant experience. If mindfulness introspection is to constitute a problem case for TT, it is not sufficient that mindfulness introspection involves thinking of properties as non-object properties. Mindfulness introspection must involve a qualitative dimension where phenomenal properties are experienced as non-object properties. Multiple philosophers have already argued that TT is implausible. Our argument is in line with arguments according to which special forms of introspective practice are in conflict with TT. Philosophers have pointed to possible conflicts between TT and oblique reflection (Loar, 2002), certain kinds of conscious meta-representation (Nicholson, 2014), an inward focus (Kind, 2003), or the phenomenology of active attention (Watzl, 2018). We point to special forms of introspection in some mindfulness practices.
We can frame the challenge posed by certain mindfulness practices by drawing on a common conception of opacity. Opacity and transparency are mutually exclusive in the following way: An experience is opaque under introspection to an individual if and only if when the individual is introspecting their experience, some properties appear to them as non-object properties. This is the conception of opacity implied by the *Mindfulness Opacity Hypothesis* (MOH). We shall not offer conclusive evidence for MOH. Our main purpose is to clarify and motivate the conditional: If MOH is true, then TT is false. We offer initial provisory evidence for MOH and sketch the kind of future work that would be needed to establish it in a more conclusive manner.

3. Mindfulness introspection

Mindfulness has its roots in Buddhist practice (Shonin *et al.*, 2015). This paper focuses on mindfulness as understood and applied in contemporary Western research and psychotherapy (for discussions of the Buddhist sources and background, see Repetti, 2022). Clinical psychologists and cognitive scientists often understand mindfulness as a metacognitive operation in which individuals intentionally increase their attention to their own present state of experience, throughout extended periods, without evaluating or affectively reacting to it (Jankowski & Holas, 2014; Bishop *et al.*, 2004; Dorjee, 2016; Wielgosz *et al.*, 2019). Due to its positive effects on mental health, mindfulness is today a core element in contemporary clinical psychology and widely researched. Long-term mindfulness practice also seems to exhibit certain biomarkers involving morphology, connectivity, and functionality of the brain (Goleman & Davidson, 2017; although, see Vago, 2022 for a critical assessment of this literature).

Many clinical psychologists have made statements that seem to strongly support MOH. They often stress the crucial difference between cognitively categorising and introspectively experiencing a mental state as a mental state:

The former [i.e. the metacognitive operation of introspectively *judging* a thought to be a mental state], even if frequently repeated, can remain isolated ‘cold’ information, whereas the latter [i.e. the metacognitive operation of *experiencing* conscious mental states as mental states], is likely to have profound effects on our views of everything, and can radically affect the way we live from moment to moment. Applied to our introductory example, this contrast highlights the difference between, on the one hand, actually experiencing thoughts as thoughts (that is, as events in the mind, rather than as direct
readouts on reality) in the moment that they occur, and, on the other hand, simply thinking about thoughts as ‘other than facts’. (Teasdale 1999, p. 147)

Clinical psychologists frequently emphasise the unique therapeutic gains of such an experiential insight (Shapiro et al., 2006).

Philosophical readers might stress that the above quote only concerns cognitive states—it refers to experiencing thoughts as mental states. By contrast, TT is often phrased as a thesis about perceptual experience, as highlighted in §2. This raises important questions about how to understand MOH. Do researchers think that only cognitive states can be opaque under mindfulness introspection, whereas perceptual states might remain transparent? Or maybe, could researchers understand the opaqueness of thoughts to be different from the opaqueness of perceptual states?

One way to think of the opaqueness of cognitive states would be to draw on the idea of cognitive phenomenology. This is the idea that cognitive states, for example conscious thoughts, have phenomenal character that cannot exhaustively be accounted for as sensory phenomenal character (Chudnoff, 2015). If one endorses this idea, one could understand mindfulness as involving introspective awareness of cognitive phenomenological properties appearing as non-object properties. If a TT-supporter is committed to strong Representationalism, they are most likely deniers of proprietary cognitive phenomenology (Tye & Wright, 2011).

MOH is not committed to the existence of a proprietary cognitive phenomenology. Assume that you deny cognitive phenomenology. Thoughts could still be introspected as opaque, namely by their involvement of sensory phenomenological properties. Take the depressive thought of “my life is hopeless”. This thought involves multiple sensory qualities like inner speech, mental imagery, or some kind of bodily-affective feeling. Introspecting this thought as opaque would then involve that these sensory qualities genuinely appeared to the subject as non-object properties. That is, the properties would not appear as properties of the objects represented by these sensory states. Hence, the conflict between MOH and TT does not presuppose that MOH accepts cognitive phenomenology. MOH is consistent with the phenomenology of thought being explained both by a proprietary cognitive phenomenology or by sensory phenomenology, as long as the phenomenal properties of the cognitive process can appear as non-object properties.

Mindfulness researchers do not only focus on mindfulness introspection of cognitive states and processes. Independently of the philosophical issues concerning cognitive
phenomenology, mindfulness researchers often describe the mindfulness introspection of perceptual states (see also the discussion of pain in §4). An example would be the influential metacognitive model of mindfulness proposed by Jankowski and Holas (2014, p. 67). The authors state that the metacognition involved in mindfulness “imposes at least two levels of cognition: (1) the lower level which refers to the qualia (basic qualities of experience such as perceptions) occurring in the present, and (2) the higher level constituted by awareness of the flowing qualia.” (Jankowski & Holas, 2014, p. 68)

Jankowski and Holas (2014, p. 68) go on to stress that mindfulness can be seen as involving a form of meta-awareness, or meta-level, at which ‘qualia are re-represented in the context of related meta-knowledge and form meta-experiences’. Meta-experiences are here experiences of experiences and appear “when a person not only thinks about the content of experience as events in the mind, but when he/she sees them as such” (Jankowski & Holas, 2014, p. 70). These authors include perceptual states as targets of this meta-awareness.

Another example is the idea of so-called dereification from the cognitive science of mindfulness (Lutz et al., 2015; Dunne, Thompson & Schooler, 2019). Dereification concerns the state in which individuals come to experience their mental states less as a manifestation of reality and more as mental states. Dereification is described as ‘experiencing thoughts as mental events, and not as the things that they seem to represent’ (Dunne, Thompson & Schooler, 2019, p. 307), and as the process where “thoughts lose their representational integrity and are experienced simply as mental events” (Lutz et al., 2015, 644). Although thoughts are again the chosen example, it is clear from the context that these authors are committed to dereification covering also perceptual states. They describe dereification as a phenomenon applying to all kinds of mental states in which ‘phenomenal content is experienced as just a mental process’ (Lutz et al., 2015, p. 644).

In general, mindfulness researchers seem to think that most, if not all, kinds of perceptual states can appear as opaque (see Jankowski & Holas, 2014; Dunne, Thompson & Schooler, 2019). However, it should be acknowledged that, as practiced in contemporary Western contexts, mindfulness often involves subjects sitting down with closed eyes and turning their attention away from the external environment to their internal states. Consequently, bodily sensations are commonly stressed as the type of perceptual states that are opaque under mindfulness (Lutz et al., 2015).

This view of mindfulness as involving a higher-level awareness of one’s own experiences as experiences—whether cognitive or perceptual in nature—is adopted by a number of psychologists and cognitive scientists studying mindfulness (e.g. Shapiro et al., 2006; Lutz et
Throughout this literature, we find researchers pointing to the difference between a form of metacognitive introspection where an individual relates merely cognitively to their first level experience and a form of introspection where an individual relates in a form of new experience to their first-level experience. The latter form of introspection is often supposed to be a special form of awareness of one’s ongoing flow of consciousness.

It is also worth stressing that a number of different notions of opacity seem to be operative in the scientific literature on mindfulness. We can capture some of this variation by distinguishing between opacity as non-object awareness (where individuals have introspective awareness of phenomenal properties that appear as non-object properties) and opacity as construction (where the content of one’s experience appear as a construction of one’s own mind). This latter notion of opacity is common in the mindfulness literature, also among philosophers (MeTzinger, 2003; Letheby, 2021). For example, cognitive scientists Lutz, Mattout, and Pagnoni (2019, p. 169-70) write:

As the meditator becomes more acquainted with the practice, he/she may acquire the skill to remain dynamically poised between the subtle anchoring to the prescribed attentional target and the increasing pull of the spontaneously arising mental content. It is tempting to see this condition as a kind of “lucid daydreaming”, optimally suited for the phenomenal “opacification” of mental processes. A mental event is said to be transparent when we have conscious access to its content, but not to its non-intentional structure or construction process. Crucially, transparency provides the phenomenal quality of being directly “in touch” with the represented entity, and is, therefore, linked to our subjective confidence in its “reality”. The opacification of mental events during meditative practice is thus equivalent to fostering their dereification, so that their provisional, constructed, dependent and ultimately impermanent nature begins to be intimately realized.

It remains an open question how opacity as non-object awareness and opacity as construction relate to each other (do they imply each other in some way or are they identical?). A TT-supporter might respond that opacity as construction seems fully compatible with TT. Introspecting one’s experience as having the properties of being “provisional, constructed, dependent and ultimately impermanent” does not have to involve that these properties appear as non-object properties. One can gain these insights through more indirect means. Subjects
might simply notice the shifting content of their experience through mindfulness, and through this they might inferentially realise that their mental states are provisional, constructed, dependent and ultimately impermanent. TT would only disallow more direct and immediate awareness of appearing non-object properties. This reply stresses a general issue. The psychologists and cognitive scientists quoted above have probably not developed their views of mindfulness on the background of the precise philosophical distinctions and discussions of TT. The above statements by Teasdale (1999), Jankowski and Holaś (2014), and Dunne, Thompson, and Schooler (2019) might therefore be too ambiguous to justify interpreting them as committing the authors to the idea of mindfulness involving opacity as non-object awareness.

We return to this issue in §5, where we discuss the relevance of experimental philosophy. For now, let us stress that although these interpretative issues do obtain, the above statements warrant acceptance of the claim that many mindfulness researchers endorse MOH. Concerning the two different notions of opacity, we interpret the relevant psychologists and cognitive scientists as stating that opacity as construction is sufficient for opacity as non-object awareness. We think this is reasonable given the general argumentation among authors like Lutz, Mattout & Pagnoni (2019). These interpretations also seem to align with how other philosophers have interpreted the relevant literature (Metzinger, 2013; Letheby, 2021).

4. Elaboration: perspective shifting, scope, and skill
In §2 and 3, we argued for the following conditional: if the Mindfulness Opacity Hypothesis (MOH) is true, then the Transparency Thesis (TT) is false. In §3, we provided evidence in favour of the antecedent. We now proceed by providing a line of philosophical interpretations of MOH. These interpretations explicate MOH philosophically in terms of perspective shifting, introspective scope, and skill. To be sure, our interpretations do not exhaust all the possibilities by which one could interpret or elaborate MOH. Future work on MOH might explore alternative routes.

We might initially ask, if experience can be opaque under mindfulness introspection, what kind of change in consciousness does mindfulness introspection involve? One answer would involve a change in perspective such that (some) phenomenological properties that prior to mindfulness introspection appeared as properties of the objects represented now appear as properties of the experience—that is, properties that appeared as object-properties transform into appearing as non-object properties. Notice that most mindfulness researchers and practitioners view mindfulness as an introspective skill (Wielgosz
et al., 2019). We expect that this transformation of the appearance of properties would demand practice and sufficient skill level in mindfulness.

What would be an example of this shift in perspective and its involved transformation? Reading the mindfulness literature, a paradigmatic example could be that of affective properties in pain experience. These properties have to do with the qualities of badness, hurtfulness, or displeasure typically involved in pain experience (Corns, 2014). The mindfulness literature contains multiple descriptions of mindfulness meditators going through a shift of perspective by which the affective properties of their pain experience transform from appearing as object-properties of the involved extramental representational object (i.e. their bodily damage) to appearing as non-object properties of their own mental states. Consider the following reports by high-level mindfulness meditators:

Pain becomes unbearable when the mind takes over saying “it is unbearable, it shouldn't happen to me, it's horrible”. Meditation [or mindfulness] showed me that the agent that distinguishes pleasant and unpleasant physical sensations, painful or not, is a mental agent, not a physical one. (Poletti et al., 2021, p. 1591)

[Mindfulness or meditation is] an opportunity… to see clearly how the mind reacts to pain. Habits become very clear. Through meditation you can see what your mind does, self-pity coming out, all worries and fears, everything comes up, and you can work with it. (Poletti et al., 2021, p. 1592)

Such introspective reports stress that mindfulness involves introspective awareness of feelings of worry, self-pity, manifestation of habits, disappointment, emotionally laden evaluation, etc. They also stress that meditators gain the introspective awareness that the negative affective properties that manifest the badness, hurtfulness, or displeasure of their pain are not properties of their bodily state but their own mind or mental state. For example, the affective properties of one’s knee pain would transform from appearing as properties of the condition of one’s knee to genuinely appearing as non-object properties of one’s own mental state. This interpretation aligns with the views of many researchers (see also Goleman & Davidson, 2017, p. 147-8).

Tying back to the issues of cognitive phenomenology in §3, one might think that this shift in perspective is only a matter of introspecting cognitive mental states as opaque. Worry, self-pity, and evaluation are not perceptual states. Yet, this is an inadequate understanding of
the present idea. The present idea is that subjects engage in mindfulness introspection of their pain experience. Pain is a complex state but it is obviously sensory or perceptual in some important sense. It is the affective properties of one’s pain experience that shift from appearing as features of one’s bodily damage to appearing as properties of one’s own mental state. It is the pain state that becomes opaque with respect to its affective properties (again, see Poletti et al., 2021; Goleman & Davidson, 2017).

One might ask now whether it is only affective properties that can be introspected opaquely through mindfulness. Mindfulness researchers seem to think that other types of properties can appear as non-object properties as well. To capture this conceptually, we might distinguish between global and local transformation. This distinction stresses that we can understand the transformation in mindfulness as varying in scope and as occurring on a spectrum. The larger the number of types of properties that can transform to appearing as non-object properties, the more global the transformation is. The more restricted the transformation is to a limited type of properties (say, limited to tactile properties), the more local it is. The degree to which a transformation is global versus local is determined by how many types of experiential properties are shifted to appearing as non-object properties.

We might view the above suggestion concerning pain experience as a more local transformation. Maybe only affective properties involved in the pain episode can appear as non-object properties, leaving it open that sensory properties still appear as properties of the represented object (the bodily damage). However, the statements from the previous §3 suggest that mindfulness researchers think that dedicated mindfulness training can enable very global transformation. For instance, Lutz and colleagues write that “thoughts [and other mental states and processes] lose their representational integrity and are experienced simply as mental events” and “phenomenal content is experienced as just a mental process” (Lutz et al. 2015: 644). These authors seem to stress that any property present in one’s awareness can appear as a non-object property (see also Dunne, Thompson & Schooler, 2019; Lutz, Mattout & Pagnoni, 2019, p. 169-70; see also Davis, 2016 for a relevant philosophical discussion).

Statements from mindfulness practitioners seem to align with this. Take again the experience of pain as an example. High-level mindfulness meditators participating in pain studies make claims like “as, what pain finally is? It is just a mental perception” (Poletti et al., 2021, p. 1592), and “pain is… the mind, it has no essence in the end. Although you [prior to the mindfulness introspection] feel it physically to a certain degree” (Poletti et al, 2021, p. 1598). One could reasonably interpret these statements as making the claim that all properties
of pain experience (e.g. location, itchiness, heat, etc.) can transform into appearing as non-object properties under mindfulness (see also Goleman & Davidson, 2017, ch. 8).\textsuperscript{18}

Mindfulness researchers and practitioners often seem to think that the more skilled a mindfulness meditator is, the broader scope of properties she can introspect as non-object properties (e.g. Lutz \textit{et al.}, 2015). Only high-level mindfulness introspectors can realise the global opacity—less skilled introspectors are restricted to more local forms. Furthermore, high-level mindfulness introspectors have greater introspective flexibility. They can fluently shift between global and local forms of opacity depending upon their goals. Consequently, opacity is an introspective accomplishment of mindfulness. This separates MOH from other challenges to TT (e.g., Loar, 2002; Kind, 2003), where opacity is not stressed as depending on one’s level of skill.\textsuperscript{19}

One way to give a philosophical spin on the claim that under mindfulness introspection phenomenal properties can be transformed from object-properties into non-object properties of one’s experience would be by using the notion of mental paint (Block, 1996). Mental paint refers to phenomenal properties that play some representational role (i.e. they enable representation of objects and their features), although these phenomenal properties are in fact non-object properties. Just like ordinary paint on a canvas enables representation of objects but is itself a non-object property of a painting, we could understand mindfulness introspection of pain as introspection of mental paint. Mindfulness meditators become introspectively aware of phenomenal properties like heat, itchiness, and negative affect that play representational roles in representing the relevant bodily damage, but under the right kind of introspection these phenomenal properties shift to appear as non-object properties. However, notice that the notion of mental paint was originally introduced to make a metaphysical argument. It was introduced to defend the claim that some phenomenal properties are non-object properties. We target TT as a phenomenological thesis and investigate whether some properties appear as non-object properties under mindfulness.

\textsuperscript{18} Despite the fact that pain experience is commonly used in objections against TT, independent of any consideration of mindfulness, many philosophers sympathetic to TT still support the claim that TT can account for pain experience (Cutter & Tye, 2011; Simon, 2019). The present idea of global or local transformation provides a novel opportunity for increasing the pressure on TT. Our version of the pain objection to TT is anchored in the scientific literature on mindfulness and the introspective reports of practitioners.

\textsuperscript{19} Mitchell (2020, p. 583) also suggests that skill levels and goals make a difference to introspective style. He draws on Lambie and Marcel’s (2002) terminology on analytic (singling out specific properties in experience) versus synthetic (holistic attention to experience) introspective awareness. In this terminology, our suggestion is that only high-level mindfulness introspectors would be able to complete synthetic introspective awareness of experiences as opaque. However, they might apply analytic introspection in relation to certain goals.
With the above in mind, future work on MOH should acknowledge the different ways in which one could philosophically elaborate MOH. Such research should also aim to compare the different elaborations. For example, Vervaeke (2022) proposes that mindfulness involves opaqueness by the breaking down of perceptual gestalts, while Letheby (2021) describes it as a matter of experiencing one’s own representations as models of one’s mind. It is not clear how these proposals relate to each other or to the interpretations presented in this section.

Let us recap. In §2, 3, and 4, we have articulated and motivated the following modus ponens:

(1) If MOH is true, then TT is false.
(2) MOH is true.
(3) TT is false.

5. Objections: introspective reports
As flagged in §3, our reasoning rests upon a particular way of interpreting statements in the mindfulness literature made by clinical psychologists, cognitive scientists, and mindfulness practitioners. Experimental philosophy would be relevant for investigating whether our interpretation is adequate, clarifying whether mindfulness researchers and practitioners really endorse MOH. Assume that MOH receives confirmation from experimental philosophy. In what position would this leave the TT-supporter? Is this confirmation sufficient to show that TT is false? The TT-supporter could object in several different ways.

Objection (1). The TT-supporter might object that no matter the result of experimental philosophy, the introspective reports of mindfulness researchers and practitioners should carry minor weight in philosophical discussions. The reason would be that these researchers and practitioners are not sufficiently trained in reflecting upon and applying the philosophical concepts. As philosophical novices, they would most likely misunderstand the concepts and misapply them to their experiences.

However, this issue of conceptual competence is exactly the kind of worry that proper work in experimental philosophy should address. Such work should handle potential cross-talk and lack of conceptual clarity, as it has attempted with other topics (Knobe et al., 2012). We need additional justification to accept that mindfulness researchers and practitioners, upon modest training, could not competently apply the concepts of transparency and opacity correctly. It is not obvious why we should be particularly pessimistic about this issue concerning MOH.
Objection (2). The TT-supporter could acknowledge that the introspective reports of mindfulness researchers and practitioners should play some role in philosophical discussions but deny that these reports put any burden of proof on the TT-supporter. The disagreement between MOH- and TT-supporters simply constitutes a stalemate with no obvious tiebreaker in sight.

This objection is problematic for a number of reasons. First, as already stated, mindfulness and the ability to introspect one’s experiences as opaque is an introspective accomplishment demanding a considerable amount of training and skill (Dunne, Thompson & Schooler, 2019; Goleman & Davidson 2017, ch. 8). Rather than an equal confrontation, one might respond that the disagreement between TT and MOH proponents is characterised by an asymmetry in introspective skills. MOH is endorsed by individuals sufficiently trained in mindfulness introspection. Second, when a philosophical view is incompatible with dominant models in the empirical sciences, the expectation is normally that the proponents of the philosophical view must either explain why the scientific models are implausible or clarify how the philosophical theory is in fact compatible with them. We should expect the TT-supporter to do the same. In sum, the onus seems to be on the TT-supporter to show that MOH is implausible.

Objection (3). The TT-supporter could stress that since introspective reports are the central evidence source for MOH, the fate of MOH is determined by the reliability of these reports. The TT-supporter could claim that one could explain the reports as simply expressing a prior bias. The TT-supporter could stress that MOH rests on canonical Buddhist sources, such as the Lankavatara Sutra. Hence, one might think that mindfulness meditators endorse MOH not because it accurately describes in an unbiased way their introspective awareness but because they are driven by a prior belief that sculpts their introspective reports.

The MOH-supporter might reply that this objection only has noteworthy force if the TT-supporter can show that we have particular reason to think that mindfulness practitioners are especially biased in their introspective reports. The mere possibility of being biased applies equally well to the TT-supporter. The reports of TT-supporters (i.e. that their experiences are transparent under introspection) might equally well reflect their prior commitment to theories like strong Representationalism. Multiple philosophers have argued that background views and broader theoretical commitments influence introspection (Martin, 2000, p. 198; Sorteriou, 2013, p. 17). Recall also the fact that mindfulness meditators are highly trained in introspection. This group performs above average in various introspective tasks, such as accurately locating
somatosensory stimuli (Fox et al., 2012). The MOH-supporter might stress that expecting this group of individuals to be especially biased is unjustified.

However, the TT-supporter might respond that it is indeed reasonable to expect the MOH-supporter to be more biased. The mindfulness practitioners and researchers that support MOH have (most often) practiced mindfulness meditation for decades. They have studied Buddhist sources that would bias them to believe in introspection of opaque experiences, and they have been members of communities that promote such beliefs. TT-supporters, on the other hand, have not been exposed to such a heavy influence.  

This is not the context to determine whether it is justified to expect MOH-supporters to be more biased by background commitments than TT-supporters. This is a central question for further research and experimental philosophy on MOH. Let us just stress that this research should also acknowledge the potential biasing of TT-supporters. TT is a cornerstone of strong Representationalism. Representationalism has often been described as the most promising route for physicalism to account for consciousness (Dretske, 1997). This does indeed seem like a strongly motivating background for endorsing TT.

The TT-supporter might formulate the present objection in slightly different terms. She might stress that the issue is not merely that MOH-supporters are biased by encounters with canonical texts or mindfulness communities. The issue is that the very instructions and ideology that subjects must accept to do mindfulness introspection demand that they believe that introspective awareness of opaque experience is possible. Hence, the problem is not that doing mindfulness exposes you to biasing sources. Rather, the problem is that to engage in mindfulness you need to accept certain beliefs about the nature of experiences, i.e. you need to accept MOH. This means that mindfulness introspection cannot be used as an introspective case to refute TT, since this kind of introspective practice presupposes antipathy against TT (see Struhl, 2022 for related discussions on meditation and the self).

It is crucial to stress that this objection only works if it can be shown that the introspective reports favouring MOH are unreliable because they express a prior commitment to the falsity of TT. The simple fact (if it is a fact) that mindfulness researchers and practitioners are biased against TT is not enough. TT is a phenomenological thesis that denies that phenomenological properties can genuinely appear as non-object properties. TT excludes introspection of opaque experiences no matter the background knowledge or motivation of subjects (Harman, 1990, p. 39; Tye, 2002, p. 139). As a phenomenological thesis, MOH is not disqualified if mindfulness

20 We are grateful to Rick Repetti for stressing this point.
practice presupposes antipathy to TT. It is only disqualified if introspective reports of opaque experience is a pure product of this antipathy instead of accurately depicting subjects’ state of introspective awareness.

6. Objection: introspective mechanisms

Another strategy would be for the TT-supporter to object that proponents of MOH must account for the introspective mechanisms that could enable mindfulness introspection of opaque experiences. Recall from §3 and 4 that researchers often think of mindfulness as involving a form of higher-level awareness. The challenge is then to elaborate the possible mechanisms of this higher-level awareness. The TT-supporter might claim that the MOH-supporter faces a dilemma. Either the MOH-supporter thinks that this higher-level awareness is established by a cognitive, thought-like mechanism, or she thinks that it is established by a perceptual, sensory-like mechanism. Yet, each choice has unattractive implications.

On the one hand, it is hard to see how a cognitive, thought-like introspective mechanism could yield the direct introspective awareness of experiences as opaque. Having a thought-like higher-level awareness of your own pain experiences will only allow you to think of this experience as an experience with non-object properties. Thinking about one’s experience as an experience is fully compatible with TT. TT is only challenged by MOH if mindfulness involves a direct form of awareness of genuinely appearing non-object properties. The only way we can make sense of such direct awareness of non-object properties is to think of it as being perceptual, sensory-like in nature.

On the other hand, adopting the view that the introspective mechanism of mindfulness is structured like a perceptual, sensory-like ability is confronted with a number of philosophical and empirical problems. The only way we can make sense of a perceptual, sensory-like introspective mechanism is to understand it as kind of internal monitoring, a form of inner sensing (as in Armstrong, 1968; Lycan, 1996). If this is indeed an implication of MOH, the

21 We only refer to contemporary, Western understandings of mindfulness. Some lines of Buddhism—like Zen and Dzogchen—often describe meditation as a non-dual state of consciousness, meaning that there is no separation or different levels of awareness. This state of consciousness is also known as pure consciousness and might involve a kind of opacity that we have not targeted directly here (Fasching, 2022; Dunne, Thompson & Schooler, 2019). We do not aim to cover these traditions in our analyses. Yet, notice that in the discussion of higher-level awareness in mindfulness, we propose that the higher-level awareness of mindfulness arises by the unfolding of an immanent reflexivity in consciousness. This might align our portrayal of contemporary, Western understandings of mindfulness with these non-dual traditions, since the relevant higher-level awareness is not portrayed as strictly separate from the first-order awareness (as in a distinct monitoring mechanism), but as an awareness arising from a latent reflexivity of consciousness.
TT-supporter would have number of ways to argue against MOH. There are at least two reasons for thinking that the idea of such an internal monitoring mechanism is mistaken.

From a more empirical perspective, there is no compelling evidence or reason to believe in the existence of this perceptual, sensory-like internal monitoring mechanism. It is hard to make sense of how and why evolution should have equipped the human brain with such a system (e.g. Carruthers, 2000, p. 210). If an internal monitoring mechanism were to be able to yield opaque introspection of first-order experiences, it would have to be able to represent the first-order experiences in all their complexity and richness. While it is easy to see the evolutionary advantage of first-order systems like the visual system with their representational and computational complexity and power, it is harder to see the advantage of a kind of internal monitoring mechanism that could match the complexity of the visual representations. While these empirical problems are important, in the present context, the philosophical issues carry more weight.

From a philosophical perspective, a number of philosophers have stressed that if we assume introspection to work by an inner sense or monitoring mechanism, this would imply that introspection involves two ontologically independent levels of states: a higher-level of states of detecting and a lower-level of detected mental states. This type of independence would allow for two possible types of radical mistakes. First, the higher-level mental state could radically misrepresent a lower-level mental state to occur when it did not (see the criticism of higher-order theory of consciousness by Neander, 1998; Farrell, 2018). Second, the higher-level mental state could misidentify the lower-level mental state as not belonging to the individual herself (Shoemaker, 1988, 1994). Authors have found these implications implausible and incompatible with our status as rational creatures (Shoemaker, 1988, 1994; Moran, 2001).

To sum up, either MOH is committed to the claim that mindfulness introspection is a thought-like mechanism or MOH is committed to the claim that mindfulness introspection is sensory-like internal monitoring mechanism. Both options are problematic. This dilemma touches upon many fundamental philosophical topics. Even though we cannot provide conclusive answers, let us outline three different ways in which a MOH-supporter could navigate the dilemma.

First, the MOH-supporter might begin by downplaying the challenge. She might stress that to get MOH off the ground, it is sufficient that the hypothesis is supported by introspective reports and researchers in the relevant fields. It is an unreasonable initial demand that MOH-supporters can provide a positive theory of the involved introspective mechanisms. This is a
demand for the later stages of MOH research. Relatedly, the MOH-supporter might stress the scope of the challenge from the opacity of mindfulness introspection. The MOH-supporter is not obliged to provide or commit to any general theory of introspection. The task is much narrower—it only concerns providing a theory of the introspection of mindfulness. This means that the MOH-supporter can be a pluralist about introspection. She could adopt other theories for more ordinary types of introspection and claim that these cases of introspection work by other mechanisms than that of mindfulness (Schwitzgebel, 2012). Rather than offering a real reply to the challenge of the dilemma, this type of response would merely deflect the challenge by downplaying its urgency and narrowing its scope.

Second, the MOH-supporter could dismiss the dilemma as resting upon the dichotomous assumption that introspective mechanisms must either be exclusively cognitive (thought-like) or perceptual (sensory-like). This assumption is controversial. Consider a theory like predictive coding (Clark, 2015; Hohwy, 2013). This theory states that the phenomenal character of any experience is created by the interaction of top-down predictive models and bottom-up sensory feedback. This also concerns introspection. The phenomenal character of introspective awareness is constituted by top-down modelling (predicting or modelling one’s state of experience) and sensory feedback updating these models (Hohwy, 2013, p. 245).

One way to elaborate MOH would be to adopt a predictive coding framework. Take the experience of pain in one’s knee. According to a predictive coding framework, the first-order pain experience in one’s knee occurs by the interaction of top-down modelling and sensory feedback. In terms of the higher-level awareness, one might think that the first-order pain experience transforms from being transparent to being opaque when an additional top-down model predicts or models the first-order pain experience itself as a mental state. This involves that (at least some of the) phenomenal properties of the first-order experience shift from appearing as object-properties of the extra-mental object (the bodily damage) to non-object properties (as properties of one’s own intra-mental state). The phenomenal character of opaqueness is an effect of the mind’s modelling of its own experience (Lutz, Mattout & Pagnoni, 2019). A predictive coding framework would allow the MOH-supporter to explain opaque experiences as involving both cognitive introspective mechanisms and perceptual introspective awareness. In this framework, mindfulness would involve a higher-order modelling of one’s own experience that resulted in direct awareness of sensory properties as appearing non-object properties.

The TT-supporter might object here. She might state this predictive coding inspired account only explains why some phenomenal properties can appear to the subject as non-object
properties. It does not explain why and how the subject is actually aware of any non-object properties. In polemic terms, this account only explains how a subject might ‘hallucinate’ non-object properties; it does not show that she actually attends to or experiences such properties under introspection.

This objection highlights an issue stressed earlier on whether to understand TT as a phenomenological or metaphysical thesis. As stressed in §1, this paper understands TT as a phenomenological thesis according to which phenomenal properties cannot appear to a subject as non-object properties. One might alternatively understand TT as a metaphysical thesis (as, for instance, Weksler, Jacobson & Bronfman, 2019). As a metaphysical thesis, TT could perhaps allow that properties can appear as non-object properties to the subject, but it necessarily denies that these appearing properties are in fact non-object properties to which the subject attends. The above predictive coding inspired account is clearly relevant to the phenomenological understanding of TT. It is beyond the present scope to evaluate how it relates to the metaphysical version of TT.

A third way to respond to the dilemma would be the following. Assume for the sake of the argument that the MOH-supporter accepts the basic structure of the dilemma and chooses the second horn according to which the higher-level awareness of mindfulness introspection is enabled by a sensory-like introspective mechanism. Still, this choice might not imply commitment to the idea of an inner monitoring mechanism.

Here is an important observation about mindfulness research. The claim that introspection is an inner form of monitoring is not widespread in the mindfulness literature. Mindfulness researchers do not typically propose any mechanistic account of inner sense. They think of mindfulness more as the act in which the mind allocates its general processing resources to its own current state of awareness by entering an explicit self-reflexive mode (e.g. Dorjee, 2016; Malinowski, 2013).

This observation suggests a way to make sense of a sensory-like introspective mechanism without committing to an inner form of monitoring. Notice that mindfulness researchers often stress that the kind of higher-level awareness involved in mindfulness can only be established by introspectively attending to one’s actually occurring first-order experience (Jankowski & Holas, 2014; Dunne, Thompson & Schooler, 2019). To establish the higher-level awareness of mindfulness, an individual would need to already be in a lower-level state of awareness, for instance, a lower-order state of pain. This state of awareness is subsequently targeted introspectively. When engaging in the higher-level awareness of mindfulness introspection, an individual always takes as their point of departure an actually occurring lower-level awareness.
One way to philosophically interpret descriptions like these would be to state that the higher-level awareness in mindfulness introspection is not established by the activation of a higher-level monitoring mechanism that is separate from the first-order experience. Instead, one might claim that the higher-level awareness of mindfulness exploits the general reflexive nature of conscious awareness. Numerous philosophers believe, for independent reasons, that consciousness is in general characterised by an immanent reflexivity, i.e. an inbuilt self-consciousness of one’s own experiential state (e.g. Frankfurt, 1988; Grünbaum, 2012). The MOH-supporter might claim that mindfulness introspection enhances or makes explicit this immanent reflexivity of consciousness to such a degree that some phenomenological properties transform from appearing as object-properties to appearing as non-object properties (see also Spackman, 2022). Mindfulness is not an introspective operation of a separate monitoring mechanism that tracks first-order mental states. It is the operations of enhancing or making explicit the already reflexive nature of consciousness.

This account stresses an important point about the unity of opaque experience in mindfulness introspection. The introspective awareness of opaque experiences in mindfulness is not a matter of bringing two separate levels of awareness together, as in synthesizing a separate first-order experience with a separate higher-level awareness. Since mindfulness is a matter of enhancing the already immanent reflexive feature of consciousness, there is no need for such synthesizing. Instead of bringing two levels of awareness together, mindfulness introspection is a matter of expanding the immanent reflexivity of an already unified consciousness of a subject.

The above proposal is only a very rough sketch and should be developed further. The important thing is that the idea of mindfulness as an unfolding of immanent reflexivity offers answers to the problems of inner monitoring outlined above. First, this idea does not commit to any idea of the human mind containing an internal monitoring mechanism for detection and representation of internal states. We can therefore dodge the demand for empirical evidence and evolutionary plausibility. Second, the proposal provides answers to the problems concerning misrepresentation and misidentification. In their traditional formulations, these problems presuppose that the relation between the higher-level and lower-level mental states is that of a causal tracking relation. If higher-level awareness of mindfulness stems from an enhancement of an immanent reflexivity of consciousness, it does not involve the same ontological separation between the different levels. The higher-level awareness is ontologically dependent upon the first-order experience since they are immanent features of the same state of consciousness. The higher-level awareness cannot occur without the first-order experience.
The above arguments are not conclusive. Yet, they seem sufficient to establish that MOH should not be discarded too quickly because of its theoretical implications about the nature of introspection.

7. Conclusion
We have provided a review of the relevant scientific literature on mindfulness, indicating that mindfulness researchers typically support what we labelled the Mindfulness Opacity Hypothesis (MOH). MOH is incompatible with the well-known philosophical Transparency Thesis (TT). We then offered a philosophical elaboration of MOH in terms of experiential shift, scope, and skill. Finally, we defended MOH against various objections. One type of objection concerned the reliability of introspective reports. We argued that the MOH-supporter is able to respond to this objection in reasonable ways, although a satisfying reply ultimately requires work in experimental philosophy. Another type of objection concerned the introspective mechanism involved in mindfulness. We argued that MOH is not committed to a functional monitoring account of introspection and leaves it open how exactly to account for the introspective mechanisms of mindfulness. Future research on MOH should develop an account of these mechanisms—we suggested some plausible directions for this work. Given the support of research on mindfulness and our theoretical considerations, we conclude that MOH is well-motivated. If MOH is true, then the Transparency Thesis is false.

References


A reductive account of mindfulness as metacognitive control

1. Introduction
Since 1966, over 16,581 academic papers have been published with the word “mindfulness” in their title, abstract, or keywords—most of them within the last ten years (Baminiwatta & Solangaarachchi, 2021). This makes the scientific literature on mindfulness rich, involving disciplines such as philosophy, psychology, psychiatry, neuroscience, and Buddhist studies (see Repetti, 2022, for a philosophical introduction). Despite this widespread research interest, one question remains unanswered: Is there a psychological capacity that is essential to mindfulness and which demarcates mindfulness from most other mental activities? This is an important research question. Identifying such a capacity would plausibly aid us in making sense of why mindfulness associates positively to phenomena such as attention control, emotion regulation, decision-making, and well-being (for a good introduction, see Goleman & Davidson, 2017). It might enable a unifying explanation of these associations: The underlying cause could be the strengthening of this particular psychological capacity. Our practical endeavors would also benefit from such explanations. Mindfulness is widely used in clinical settings. Knowing the nature of this essential and demarcating psychological capacity could aid us in designing efficient mindfulness protocols and in instructing service users in the best possible way (Garland, 2007; Dorjee, 2016).

Today, many researchers believe that the essential and demarcating psychological capacity of mindfulness is a form of metacognition or, more precisely, a form of metacognitive control (e.g., Jankowski & Holas, 2014; Dorjee, 2016; see also philosophical discussions, such as Repetti, 2018; Zawidski, 2019; Kachru, 2022, p. 100; Timalsina, 2022, p. 314). The present paper targets this widespread idea and provides a brief review of the relevant literature (§2). The paper has a negative and a positive aim. First, I argue that the current accounts of how to understand mindfulness as form of metacognitive control fail. This concerns various issues relating to conceptual clarity and the ability to distinguish metacognitive control from cognitive control (§3). Second, I propose a novel account of metacognitive control by introducing the idea of metacognitive goals. I argue that this account adequately describes the core psychological workings involved in mindfulness. The account also clarifies why metacognitive control is only a necessary and not a sufficient condition for mindfulness (§4). Finally, I argue that the account motivates two novel theses on the nature of mindfulness, namely that we can
give a reductive explanation of the metacognitive control of mindfulness and that this control is a distinct kind of mental action (§5).

2. Mindfulness

Imagine Lucy, who suffers from depression. Lucy enrolls in a mindfulness program with the aim of therapeutic effects. Here, she practices the two paradigmatic mindfulness techniques: focused attention meditation and open monitoring meditation (Lutz et al., 2008). Under focused attention meditation, Lucy is to focus her attention on her own breath. This involves her remaining aware of her own attentional focus and deliberatively bringing it back to her breath when it has wandered. Under open monitoring meditation, Lucy must monitor her own ongoing thoughts, feelings, sensory impressions, etc., without being immersed in them. She must remain detached. Such detachment depends upon Lucy holding certain second-order attitudes toward her own mental processes. These are attitudes of nonreactivity (i.e., not acting or interfering with the processes) and disidentification (i.e., viewing the processes from a third-person observational perspective as transient states that are not expressive of her own identity).

Contemporary researchers seek to provide a scientific account of mindfulness. They seek to clarify the psychological, neural, bodily, and social workings of this practice. At the center of this endeavor is the question: Is there a psychological capacity that is essential to mindfulness and which demarcates mindfulness from most other mental activities? Essential means this capacity would constitute the core functioning of mindfulness. Demarcating means this capacity would be unique to mindfulness or at least not commonly involved in most other operations. It would be of great interest to identify such a psychological capacity. Mindfulness is a core element in some of the most efficient contemporary cognitive psychotherapies, so-called third-wave cognitive-behavioral therapies (Hayes et al., 2011). Further, mindfulness training is associated with an increase in some aspects of cognitive control (Gallant, 2016), and extensive practice of mindfulness plausibly involves changes to consciousness that are both philosophically and scientifically puzzling (Lange & Grünbaum, 2023). Explanations of these

---

22 Some readers might think that only open monitoring meditation is a form of mindfulness, not focused attention meditation. I follow standard psychological theorizing by including both meditation forms as mindfulness (see Segal et al., 2018). §4.2 discusses whether compassion meditation practices should be included.

23 Lange (2023) proposes to understand these attitudes as stances, i.e., complex clusters of propositional attitudes, aspects of control, and commitments to certain plans and values. I shall not go into detail about this matter here.
effects and associations would benefit substantially from knowing if there is an essential and demarcating psychological capacity at play.

In search of an answer, one might consult the original Buddhist sources. Many Buddhist sources consider mindfulness a form of remembering. To be mindful is to have a form of retention of the Buddhist teachings and to approach the relevant situation with these in mind (Gethin, 2011). The Buddhist tradition contains a wide and rich supply of theories that unpack the more exact mental faculties involved in such remembering. Yet, one might worry that this notion of mindfulness is too laden with Buddhist ideas on the nature of the mind. Arguably, this approach is simply too anchored in the Buddhist tradition to provide a scientifically “sober” account of the essential and demarcating psychological capacity of mindfulness (notably, this relates to controversial ideas on the metaphysics of consciousness and the self; see Thompson, 2022).

Instead, researchers might consult more secular descriptions of mindfulness. Such descriptions appear from the 1950s onward. One prominent proposal is that mindfulness is intentional awareness of one’s own present experience in a nonjudgmental way (Kornfield, 1977; Kabat-Zinn, 1990; see also Dreyfus, 2013). Authors believe that this approach does not assume any prior religious or spiritual commitments (although, see Ratnayake & Merry, 2018). Yet, even if one agrees, this description seems insufficient in providing the desired scientific picture of mindfulness. The description does not inform us about the kind of psychological capacity that underlies this awareness. If we are to understand mindfulness and its relation to other scientific topics—such as attention, agency, consciousness, and perception—it is highly desirable to understand the nature of the underlying psychological capacity in more precise scientific terms (Dorjee, 2016).

2.1 Metacognitive control
Today, the most prominent proposal for conceptualizing mindfulness in scientific terms is that the essential and demarcating psychological capacity of mindfulness is a form of metacognition or, more precisely, a form of metacognitive control. Authors in psychology (Jankowski & Holas, 2014), cognitive neuroscience (Dorjee, 2016) and philosophy (Repetti, 2018; Zawidski, 2019; Timalsina, 2022) endorse or sympathize with this idea. Three interrelated reasons motivate the proposal.

First, the topic of metacognition and metacognitive control is well known in other scientific literature, such as that of educational psychology (Hacker et al., 2009) and cognitive neuroscience (Fleming & Lau, 2014). Hence, approaching mindfulness as an exercise of
metacognitive control seems attractive, since it approaches a seemingly esoteric practice (i.e., mindfulness) through a standard scientific notion (i.e., metacognition). One could hope that this approach would enable us to bridge the mindfulness literature with other scientific domains. We might draw on other metacognition research to investigate the nature of mindfulness and vice versa.

Second, the notion of metacognitive control might enable us to make sense of the core characteristic of mindfulness, namely that it is a very explicit form of control of one’s own mental processes (Garland, 2007). This explicitness concerns both self-awareness and self-regulation. Recall Lucy’s mindfulness training above. In terms of explicit self-awareness, Lucy is actively introspecting her own relevant mental processes (either her attentional focus or the continuous unfolding of mental processes). In terms of explicit self-regulation, Lucy is also actively regulating her own relevant mental processes (either to focus on her own breath or to have certain attitudes towards her own processes). Most everyday undertakings do not involve such explicit self-awareness or self-regulation. Take the example of Lucy searching for a coffee shop on a busy shopping street. Here, Lucy will not actively introspect her own mental processes—instead, her awareness will be oriented toward her surroundings. Further, instead of explicitly regulating her own mental processes, Lucy will fluently regulate her own mental and bodily processes in interaction with the environment. If Lucy explicitly regulates something, it is her position within the environment (avoiding other pedestrians, opening the door to the shop, etc.). Authors propose that the notion of metacognition is the most appropriate way to adequately describe this explicitness in scientific terms (e.g., Teasdale, 1999; Dorjee, 2016; Zawidski, 2019). This makes sense. Researchers generally apply the notion of metacognition to refer to self-monitoring and self-regulation (Proust, 2013).

Third, the notion of metacognitive control enables us to explain the relationship between mindfulness and more ordinary cognitive control. Notice that it is not obvious that there exists one standard definition of cognitive control. Yet, we might say that cognitive control concerns the ability of agents to structure their own mental processes (perception, thinking, planning, etc.) in accordance with their relevant goals (Altman & Grey, 2008). Research suggests that exercising mindfulness involves execution of cognitive control, and extended practice of mindfulness seems to support core aspects of cognitive control (Holas & Jankowski, 2013;

---

24 Some philosophers believe that attention control always involves a kind of self-awareness, i.e., when agents control their attention, they are also aware of the focus of their attention. However, these philosophers take such self-awareness to be implicit (e.g., Metzinger, 2013; Watzl, 2017). It is a matter of having an inbuilt awareness of one’s own attention while controlling it, not a matter of taking one’s own attention as a direct object in awareness.
Nevertheless, researchers also stress that we cannot exhaustively describe mindfulness merely as agents executing cognitive control (Jankowski & Holas, 2014). One way to stress the difference is in terms of explicitness, as mentioned above: Mindfulness is a more explicit control of one’s own mental processes than ordinary cognitive control. Another way to stress the difference is in terms of orientation. Ordinary cognitive control is most often more world-oriented; it involves awareness and regulation with respect to external matters (as in Lucy’s search for coffee). Mindfulness is more self- and mind-oriented; it involves explicit self-awareness and self-regulation of one’s own mental processes (as in Lucy’s two mindfulness meditations). This makes the relation twofold: Mindfulness and cognitive control are intimately related, yet they are also different in some important sense. Conceptualizing mindfulness as metacognitive control might enable us to make sense of this relation (Repetti, 2018; Zawidski, 2019; Jankowski & Holas, 2014).

3. Preceding accounts
Yet, if the idea that mindfulness is a form of metacognitive control is to have any bite, we must have a more elaborated account of the nature of this metacognitive control. The mindfulness literature contains multiple proposals on definitions. One proposal is very straightforward. Some psychologists seemingly characterize metacognitive control implicitly as the control of one’s own mental processes (one could interpret Wells, 2019, and Brick et al., 2016, in this way). In philosophical terms, they seemingly endorse the following definition:

(1) An agent, A, exercises metacognitive control of mental process, M, if and only if A controls M.

Definition (1) is clearly unsatisfying, but it is important to show why. The definiens is uninformative. It does not provide any insights into the distinct nature of metacognitive control. It does not make sense of why metacognitive control is particularly explicit, nor why it is distinct from ordinary cognitive control. It also deflates the concept of metacognitive control. According to (1), Lucy would exercise metacognitive control when she controls her perception in search of a coffee shop. We do not want to categorize such cases as instances of metacognitive control, but merely as exercises of cognitive control.

Another proposal is to define metacognitive control by its involvement of a special type of knowledge, namely metacognitive knowledge (authors like Jankowski & Holas, 2014, and
Efklides, 2008, 2011, seem sympathetic to this approach. We can outline this proposal as follows:

(2) An agent, A, exercises metacognitive control of mental process, M, if and only if A controls M by utilizing her relevant metacognitive knowledge.

To assess (2), one must offer an account of what exactly metacognitive knowledge is, including an elaboration of what it means to apply such knowledge. Researchers often characterize metacognitive knowledge as the knowledge agents have about their own mental processes (Chen & McDunn, 2022). They sometimes divide this knowledge into various types, such as propositional (involving propositional knowledge about one’s own mental processes), procedural (involving know-how about one’s own mental processes), and strategic metacognitive knowledge (involving “if-then” knowledge about how to mentally respond to certain cues). Despite this typology, crucial aspects remain unclear with regards to definition (2). Many questions remain unanswered, such as: Is metacognitive knowledge always accessible or reportable for the agent who possesses it?; When an agent applies metacognitive knowledge, does this always involve her having awareness of doing so?; Does procedural and strategic metacognitive knowledge reduce to propositional metacognitive knowledge? (as intellectualists generally argue, Stanley & Williamson, 2001). I do not pose these merely as rhetorical questions. They stress the requirements for elaborating definition (2). This is no trivial task. Currently, we have no developed account of metacognitive knowledge that enables a satisfying answer to these issues.

Considering the problems facing definitions (1) and (2), one might argue that the only way to account for metacognitive control is from a first-person perspective, i.e., from the perspective of the agent exercising this form of control. Some authors seem to apply this strategy. They define and characterize the metacognition or metacognitive control of mindfulness by aspects such as the awareness and intentionality it involves (e.g., Teasdale, 1999; see also Yates et al., 2017). In fact, one reasonable interpretation might be that these authors use the idea of explicitness as the definiens of metacognitive control, as follows:

(3) An agent, A, exercises metacognitive control of mental process, M, if and only if A’s control of M involves A having explicit self-awareness and self-regulation of M.
I do not think that we should accept this approach. One problem is that definition (3) does not make sense of why metacognitive control is a particularly explicit form of control. It simply states that it is. Supporters of definition (3) might argue that this is unproblematic. They might stress that since explicit self-awareness and self-regulation are not common features of our ordinary control operations (e.g., searching for a coffee shop), these features are sufficiently characteristic to demarcate metacognitive control. This response is problematic, at least if we follow common philosophical standards. We should be parsimonious when it comes to psychological concepts in science. We should only accept a concept into our repertoire if we can provide a working definition of it and demonstrate that the concept is important for some theoretical or explanatory purpose. In fact, numerous philosophers have expressed skepticism toward concepts concerning metacognition specifically and have stressed that these concepts do not fulfil such requirements (see critical discussions of metacognitive feelings and procedural metacognition, Carruthers, 2017; Smith, 2014; for a defense, see Proust, 2014). Considering this economical approach and skepticism, definition (3) does not seem sufficient for accepting the concept of metacognitive control into our repertoire of psychological concepts. It is hard to see what kind of important theoretical or explanatory benefit the definition provides.

Finally, consider recent discussions on mindfulness in contemporary philosophy of mind and philosophy of psychiatry. According to Zawidski (2019), mindfulness trains a specific metacognitive skill. Roughly, this is the skill of regulating one’s own mental processes so that they support one’s well-being. Relatedly, Leder and Zawidski (2023) have recently argued that psychological disorders and illnesses—ranging from mood disorders to psychotic conditions—are generally caused by a dysfunction of this skill. This is a very ambitious claim, which we shall not evaluate here. Instead, let us investigate whether Zawidski (2019, 2021) and Leder and Zawidski (2023) provide an attractive route for a definition of metacognitive control. I think the most promising aspect concerns a specific distinction drawn in Leder and Zawidski (2023, p. 9). The authors distinguish metacognitive skill by stating that this control directly targets mental processes, while other forms of skill and control merely make use of mental processes. It is not totally obvious what the authors mean by this distinction. I think the strongest interpretation is something like the following. Take the control involved in perception-motor skills (such as those of a tennis player) or intellectual skills (such as those of a chess player). Leder and Zawidski (2023) seemingly claim that even though such instances of control obviously involve regulation of mental processes, this regulation is only instrumental in attaining other goals. For example, a tennis player controls her perceptual processes...
instrumentally to win a point, as a chess player controls her cognition instrumentally to succeed in a strategic maneuver on the board. By contrast, metacognitive control targets mental processes directly: It regulates mental processes as ends in themselves. For example, when Lucy regulates her own attentional focus on her breath, she does not do this for other instrumental purposes involving the external environment. The relevant focus of attention is an end in itself. Hereby, Leder and Zawidski (2023) seem to endorse the following:

(4) An agent, A, exercises metacognitive control of a mental process or state, M, if and only if A’s control of M is not instrumental to other endeavors.

I think (4) gets something right about the metacognitive control of mindfulness, namely that metacognitive control targets mental processes as goals in themselves. I return to this aspect in §4. However, I do not think that definition (4) succeeds. One major problem is that (4) is chauvinistic: It excludes various instances of mindfulness as exercises of metacognitive control. Recall the example of Lucy and her depression. Lucy enrolls in a mindfulness protocol to treat her depression. This means that when Lucy regulates her own attentional focus (on her breath) or second-order attitudes (of nonreactivity or disidentification) she does so with the underlying aim or instrumental purpose of changing her depression. The psychological literature contains related examples. Athletes frequently deploy mindfulness to enhance their performance. Mindfulness enables control of attentional focus, and second-order attitudes can be helpful in managing performance anxiety (Brick et al., 2016). Here, an athlete controls her attention or higher-order attitudes with the instrumental purpose of performing better. Moreover, studies in social psychology show that mindfulness plausibly decreases certain implicit biases (e.g., Lueke & Gibson, 2015). Hence, an agent might exercise mindfulness with the instrumental purpose of decreasing her implicit bias and improving her epistemic performance. Definition (4) excludes these target cases of mindfulness as instances of metacognitive control. This is unacceptable—the job of the definition is precisely to include these as occurrences of metacognitive control.

4. Metacognitive goals
Considering these failures, I now present a novel proposal. This proposal defines metacognitive control in terms of the goals that this control involves. Let me begin with a simple observation. Recall Lucy’s two mindfulness exercises: focused attention and open monitoring meditation. When Lucy is doing these exercises, she is aiming to succeed in something. That is, she is
aiming to succeed in focusing her attention on her breath or on establishing certain second-order attitudes. I think we should conceptualize this as Lucy having certain goals. We could call these metacognitive goals, since they are goals about one’s own mental processes. That is, Lucy has metacognitive goals about her focus of attention and her second-order attitudes. My claim is that we can conceptualize Lucy’s mindfulness exercises adequately as activities in which Lucy implements certain metacognitive goals. First, consider what we shall label first-order attentional metacognitive goals. We shall define these as follows:

A goal, G, of an agent, A, is a first-order attentional metacognitive goal if and only if G’s content concerns the focus of A’s first-order attention.

I define first-order attention negatively here. Roughly, first-order attention is attention toward objects or processes that are not themselves instances of one’s own attention. Hence, I take instances of first-order attention to be those of attending to other pedestrians, the bicycles and cars on the road, or the letters on the computer screen. Further, I take instances of attention that do not qualify as first-order attention to be those of attending to one’s own visual gaze or one’s own line of thinking. Such a gaze or line of thinking are themselves instances of one’s own attention. Some readers might stress that attention is a philosophically controversial topic. What counts as an instance of first-order attention might depend upon commitments to broader theories on the nature of attention (Mole, 2011). This might be correct, although I expect that these theories converge in most cases. It is beyond the present scope to discuss this further, and it is not crucial in assessing the central arguments of the present paper.

Second, consider what we shall label second-order attitudinal metacognitive goals. We shall define these as follows:

A goal, G, of an agent, A, is a second-order attitudinal metacognitive goal if and only if G’s content concerns A’s second-order attitudes.

I define second-order attitudes as those attitudes that agents can have toward their own mental processes. There are many examples of such second-order attitudes. For example, an agent might have an attitude of approval toward some specific desire (e.g., the agent approves of her desire to go for a run), or she might have an attitude of disapproval toward some other desire (e.g., she might disapprove of her desire to have coffee after 5 p.m.). Another example would be an agent having a certain attitude toward certain beliefs. For example, depressive
patients often consider their depressive rumination to be an accurate readout of reality (Papageorgiou & Wells, 2001). A central element in certain therapies is to change these attitudes so that patients have a more skeptical attitude toward the accuracy of these ruminative thoughts.25

I think this idea of metacognitive goals provides an attractive path to defining the metacognitive control of mindfulness. We can conceptualize Lucy’s focused attention meditation as the activity of implementing a first-order attentional metacognitive goal. Lucy has the goal of focusing her attention on her own breath, which she implements. Further, we can conceptualize Lucy’s open monitoring meditation as the activity of implementing second-order attitudinal metacognitive goals. Recall that open monitoring meditation involves the attitudes of disidentification and nonreactivity. These are second-order attitudes that target one’s own occurring first-order process, enabling detachment. Lucy is implementing these goals. These conceptualizations offer a novel definition of the metacognitive control of mindfulness, namely:

(5) An agent, A, exercises metacognitive control of a mental process, M, if and only if A controls M through a first-order attentional or second-order attitudinal metacognitive goal.

That is, in focused attention meditation, Lucy exercises metacognitive control because she controls her first-order attention through her first-order attentional metacognitive goal concerning focus on her own breath. When doing open monitoring meditation, Lucy is exercising metacognitive control because she controls her higher-order attitudes through her higher-order attitudinal metacognitive goal of having the second-order attitudes of disidentification or nonreactivity. As I shall discuss further below, such control seems substantially different from the control we exercise in most of our ordinary actions. Take the goal of finding a coffee shop, driving home, or finishing a paper. We would normally explain

25 Let me stress that this notion of second-order attitudes is neutral with respect to broader philosophical issues on introspective awareness, for example the transparency thesis. This thesis claims that one cannot introspect intrinsic properties of one’s own mental states—one can only introspect representational objects or properties of these (Tye, 2002). Even though one might claim that mindfulness does challenge this thesis (Lange & Grünbaum, 2023), the present idea of second-order attitudes does not imply such a position. One can think of second-order attitudes as in accordance with the transparency thesis. Having a second-order attitude toward some mental state does not need to imply that you have nontransparent or opaque awareness of this state. For example, if an agent disapproves of her desire for a cup of coffee, then she might only be aware of representational properties of this state (e.g., the liquid of the coffee, the cup, certain bodily sensations, affects, inner speech utterances, movements, and quasi-sensory simulations).
these actions as agents implementing goals that concern the external environment, notably certain objects in the environment and the agent’s relation to these objects. It is important to stress that these goals, with content about external matters, obviously do determine the agent’s mental processes. Having the goal of finding a coffee shop does determine my perception, thinking, and motoric activity. However, the content of the goal does not concern these mental processes. It concerns the external environment and my position within it.

Now, some philosophical readers might be skeptical to the proposal of (5). They might stress that the notion of metacognitive goals is not sufficiently clear. One major issue could be that the notion makes controversial assumptions with respect to the philosophical literature on goals. One assumption could concern the imperative force of goals. Philosophers disagree about how goals direct or demand behavior (Schroeder, 2020). Some philosophers believe that goals get their imperative force solely due to their special kind of imperative content. Other philosophers believe that the imperative force of goals is not in their content but in certain attitudes toward content. What position does the idea of metacognitive goals commit to? No specific one, would be my answer. The above notion of metacognitive goals is compatible with both positions. This notion only makes the claim that the content of goals can concern one’s own attentional focus or second-order attitudes. It is silent with respect to whether this content is of a special imperative format or whether these goals get their imperative force from the involvement of certain attitudes.

Other readers might stress the philosophical debates about the format of the content of goals. Many philosophers believe that the content of goals can be of both propositional format (where the content can be expressed in propositions) and non-propositional format (where the content cannot be expressed in propositions but is encoded in other formats, such as sensorimotor ones: Pacherie, 2008). Relatedly, philosophers often draw another (orthogonal) distinction between personal-level goals (which agents can access and report) and sub-personal level goals (which agents cannot access and report: Christensen et al., 2016). Again, I do not see any demand to commit to certain positions on these issues. Yet, in terms of providing an account of the metacognitive control of mindfulness, I consider it appropriate to understand these metacognitive goals as personal-level goals in a propositional format. The following sections operate with this interpretation.

Let me then provide five reasons for why the definition of (5) is more attractive than the preceding definitions reviewed in §3. First, unlike definition (1), definition (5) is informative and does not deflate the concept of metacognitive control. There are many instances where definition (5) would not be satisfied. Searching for a coffee shop, driving a bicycle, or finishing
a paper are not activities where we would ascribe first-order attentional or higher-order attitudinal metacognitive goals to agents (recall that these metacognitive goals are personal-level goals with propositional content, not simply sub-personal goals).

Second, definition (5) does not operate with an obscure concept at its core, as definition (2) does. Definition (2) demarcates metacognitive control as the application of metacognitive knowledge, although we have no account of the nature of such knowledge and what it means to apply it. The notion of metacognitive goals, as in first-order attentional or second-order attitudinal metacognitive goals, seems more comprehensible. These are a type of goal distinguished by their content. This account appears clearer and more promising than providing a theory of metacognitive knowledge and its application.

Third, definition (5) can make sense of the explicitness of metacognitive control, i.e., the involvement of explicit self-awareness and self-regulation. One attractive way to do this is by drawing on Irving’s (2021) general idea that goals are phenomenological frames. Irving (2021) argues that our goals carve out our awareness into relevant and distractive phenomena. For example, if an agent has the goal of finding a coffee shop, this goal will frame her awareness into relevant phenomena (e.g., shop signs, objects like coffee cups, certain smells) and distractive phenomena (e.g., a plane passing over her, someone talking on her cell phone). Being a type of goal, metacognitive goals are also phenomenological frames. Nonetheless, since metacognitive goals do not concern the external environment but rather our own mental processes, the framing is not a matter of carving the external environment into relevant and distractive phenomena. Instead, metacognitive goals frame our awareness so that the mental process(es), which the goal concerns, is the relevant phenomenon, while other phenomena are distractive. For example, the first-order attentional metacognitive goal of focusing your attention on your own breath carves out your awareness so that it is your attentional orientation and your own breath that are the relevant phenomena. Other phenomena are distractive, involving other mental states (e.g., spontaneous thoughts about what you are having for dinner) or external objects (e.g., the sounds of the busy shopping street outside your window). This provides a simple explanation of why metacognitive control involves explicit self-awareness and self-regulation. This is in contrast to definition (3), which simply uses explicitness in its definiens.

Fourth, definition (5) can distinguish between cases of cognitive control and metacognitive control. The general divide is that these are distinct forms of control because they involve distinct forms of goals. The metacognitive control of mindfulness involves first-order attentional or second-order attitudinal metacognitive goals, which our ordinary exercises
of cognitive control do not. Some readers might be skeptical about this distinction. They might stress that some exercises of cognitive control do not involve goals related to our external environment but rather goals about our own mental processes. An example might be the cognitive control involved in explorative thinking or memory retrieval. This objection is relevant. I will return to it in §4.1. For now, the important thing is that definition (5) proposes that the major divide between ordinary exercises of cognitive control and the metacognitive control of mindfulness is a matter of the goals they involve. This distinction is clear, and I will argue that it is also satisfying.

Fifth, definition (5) avoids the problem of chauvinism relating to definition (4). Recall that definition (4) states that when subjects execute metacognitive control, they regulate their own mental processes without this being instrumental to other purposes. This was too demanding, since it excluded therapeutic or performance-oriented uses of mindfulness. Definition (5) can avoid this problem by drawing on standard theory in philosophy of action. This concerns goal hierarchies. Philosophers commonly think that an agent’s goals can be of different levels and that these levels interact in action control. Again, take the example of navigating a busy shopping street to find a coffee shop. Here, the agent has the overall goal of finding the coffee shop, sometimes called an ultimate or strategic goal. Moreover, she also has a more specified goal of navigating the acute landscape of the crowd, sometimes called a proximal or situational goal. Lastly, she also has very fine-grained goals of making small bodily changes in this landscape of people and obstacles, sometimes called a motor or an implementation goal (Pacherie, 2008; Christensen et al., 2016).

I think we should explain definition (5) further by stating that metacognitive goals are goals at the more proximal level. Proximal goals are accessible to the agent, in accordance with the previous specification of metacognitive goals as personal-level goals. The important thing is that proximal goals can be coupled with various ultimate goals. As Lucy, one can have a certain proximal metacognitive goal of focusing one’s attention on one’s own breath, yet this happens on the background of the ultimate goal of therapeutic effects. Moreover, athletes might exercise mindfulness with ultimate goals about perception-motor performance, while other agents might exercise mindfulness with ultimate goals about bettering epistemic performance. If we elaborate definition (5) on these lines, we can allow that these are cases of metacognitive control, even though they involve instrumental control of one’s own mental processes. As a result, we should update our definition:
An agent, A, exercises metacognitive control of a mental process or state, M, if and only if A controls M on the basis of either a first-order attentional or a higher-order attitudinal metacognitive goal \textit{at the proximal level.}

4.1 Scope of metacognitive control

Let me turn to discussing two interrelated worries concerning scope. Some readers might worry that (5)* is too liberal in the sense that it categorizes certain operations as metacognitive control when this is inappropriate. Imagine that you make a cup of coffee with the purpose of remaining focused on writing your paper. We can conceptualize this purpose as involving the first-order attentional metacognitive goal of focusing your attention on your paper. Further, we can say that this metacognitive goal is involved in the control of your attention, perception, and motor activity. It is on the basis of this metacognitive goal that you perceive, plan, and move to make your coffee. Therefore, following definition (5)*, we have a case of metacognitive control: You control your relevant mental processes in making a cup of coffee by the first-order attentional metacognitive goal of focusing on your paper. Yet, this seems mistaken. We would say that your activity is merely an exercise of cognitive control.

However, I think we can handle this alleged counter-case by stressing the demand that metacognitive goals must be involved at the proximal level of an operation. This is not the case with the above activity of making a cup of coffee. The metacognitive goal of staying focused on your paper is \textit{not} the proximal goal that determines your perception, cognition, and motor activity. The proximal goal determining these activities is the goal of making a cup of coffee, i.e., a goal concerning the external environment. The metacognitive control is the more ultimate goal in the background. Hence, it is not a case of metacognitive control according to definition (5)*.

Skeptics might present another case, then (relating to a point flagged in §4 above). Consider the case where you want to have dinner at a restaurant, though you are not sure exactly which one. You form the goal of coming up with a good idea about a specific restaurant. We can call this a metacognitive goal, i.e., a goal about generating a mental representation of a certain kind. You engage in some explorative thinking to generate this idea. Seemingly, we have a case of metacognitive control. A metacognitive goal functions on the proximal level in an agent’s control of her mental processes (her thinking, simulation, memory, imagination, etc.). Likewise, consider another case. You are standing at the market to buy groceries, yet you have forgotten your list of items that you needed to get. You try to recall the items on the list. We can also call this a metacognitive goal, i.e., a goal concerning mental representation of
certain objects. Further, it functions on the proximal level in controlling your mental processes. Again, we have a case of metacognitive control, according to (5)*. This is problematic. These cases are simply cases of cognitive control.

I think the first response to these objections would be that they misread definition (5)*. Definition (5)* does not state: An agent, A, exercises metacognitive control of a mental process or state, M, if and only if A controls M on the basis of a metacognitive goal at the proximal level. Instead, definition (5)* states that the involved metacognitive goals must be of a certain kind, namely first-order attentional or second-order attitudinal metacognitive goals. With this in mind, it comes down to whether we are willing to ascribe first-order attentional or second-order attitudinal metacognitive goals to the agents in the two cases. I am not sure this would be appropriate. The metacognitive goals that drive the mental operations seem to be of a different kind here. The core activities in these two cases are not those of focusing attention or establishing certain second-order attitudes. The core activities are those of forming mental representations with certain practical or epistemic virtues—that is, settling on a restaurant (with the best combination of food, feasibility, service, etc.) or accurately representing the objects on the list. This seems importantly different from first-order attentional or second-order attitudinal metacognitive goals.

Skeptics might respond that this reply reveals that definition (5)* is arbitrary. Why exactly is it first-order attentional and second-order attitudinal metacognitive goals that demarcate metacognitive control, and not simply any kind of goal we could call a metacognitive goal (i.e., any kind of goal that concerns one’s own mental processes, for example mental representation of restaurants or grocery items)? Here, I would like to remind the reader of the present discussion and purpose. This paper aims to provide an account of the specific form of metacognitive control relevant to mindfulness. My claim is that this specific control involves these two specific kinds of metacognitive goals. This does not make definition (5)* arbitrary. It simply stresses the ambition and focus of (5)*.

This being said, I acknowledge that definition (5)* might come with gray-zone cases, where it is hard to determine whether or not an operation qualifies as metacognitive control. The central issue will always be whether we are willing to ascribe first-order attentional or second-order attitudinal metacognitive goals, at the proximal level, to the relevant agent and
her operation. Discussing the exact criteria for such ascription is obviously of relevance to the proposal of (5)*, yet it is beyond the present scope to debate this issue further.26

Another objection would be that definition (5)* is too narrow because it neglects other uses of the concept of “metacognitive control” in other literature than that of mindfulness. Literature like that of neurodegeneration, motor neuroscience, and philosophy of action also use the concept of “metacognitive control” (e.g., Souchay & Isingrini, 2004; Pacherie & Mylopoulos, 2021). In fact, this relates back to the above two cases on the explorative thinking of restaurants or the recalling of grocery items. Perhaps it is not a virtue if definition (5)* excludes these cases as occurrences of metacognitive control. For example, Fox and Christoff (2014) argue that a form of metacognitive regulation facilitates and controls explorative thinking and mental simulations, like those of spontaneous thoughts or mind-wandering (for example, about restaurants). Roughly, such regulation concerns the ability of agents to intentionally and knowingly direct their capacities of creating mental content. Likewise, Proust (2013) argues that many epistemic operations, like the one of recalling the grocery items, relies upon certain dynamics of metacognitive monitoring and regulation. Following these independent theories of metacognition, it might make sense to label such cases of explorative thinking and recalling of objects as instances of metacognitive control. Hence, definition (5)* is, actually, too narrow.

However, I think this objection demands too much of the present context. This paper does not have the ambition of offering a concept of metacognitive control across a range of literature. Yet, if this were the ambition, I would actually be optimistic in (5)* succeeding or, at least, providing substantial resources. The idea of metacognitive goals seems to fit nicely into various other literature. For example, the notion of metacognitive control in educational and sports psychology concerns agents regulating their attentional focus or mental strategies in solving tasks and performance pressure (Brick et al., 2016). There might be a good chance that

---

26 Another worry might be that definition (5)* categorizes certain attention tasks, as those in visual attention studies, as instances of metacognitive control. In some of these tasks, agents are to attend to certain symbols or properties (e.g., the “Stroop task,” MacLeod, 1992). Hence, one might think that such tasks involve first-order attentional metacognitive goals about one’s own focus (e.g., focus on the color of the letter and not its linguistic content). The challenge is that we view these tasks as measuring cognitive control, not a special form of metacognitive control. Yet, I do not think we can handle such a challenge by stressing we should not ascribe first-order attentional metacognitive goals at a proximal level to agents in these attention tasks (again, recall that these are personal-level propositional goals). Tasks like the Stroop task concern agents attending to certain objects and properties to report these accurately and quickly. I think this task of reporting is the relevant proximal goal of the agent. Such a task obviously involves control of attention, but the control is not by a proximal first-order attentional metacognitive goal. On the contrary, if the task were simply to attend to certain objects or properties on the screen, with no task of reporting these, then this would be a first-order attentional metacognitive goal—hence qualifying as metacognitive control.
definition (5)*, and its idea of metacognitive goals, would adequately describe metacognitive control in this psychological literature as well. As with the other literature mentioned above, we should at least remain agnostic.

4.3 Scope of mindfulness

Another worry would be that it is not clear what the exact relation between the concept of metacognitive control in definition (5)* and mindfulness is. We can think of three basic logical relationships.

* Necessary and sufficient. An agent, A, exercises mindfulness if and only if A exercises metacognitive control as defined by (5)*.

* Sufficient. If A exercises metacognitive control as defined by (5)*, then A exercises mindfulness.

* Necessary. An agent, A, exercises mindfulness only if A exercises metacognitive control as defined by (5)*.

I do not think we should endorse “Sufficient” nor “Necessary and sufficient.” The basic reason is that we might come up with cases that qualify as metacognitive control under (5)*, yet they are not occurrences of mindfulness. For example, if we accept that cases such as explorative thinking or the recalling of objects qualify as metacognitive control, then we would still *not* say that they qualify as mindfulness. We can come up with other cases as well. Imagine an agent that implements the second-order attitudinal metacognitive goal of relating to her own thoughts as if they were induced by evil agents from the outside. This would qualify as metacognitive control, yet it is obviously not an instance of mindfulness. We can, hereby, conclude that metacognitive control, as defined by (5)*, is *not* sufficient for mindfulness.

Instead, I think we should endorse “Necessary.” It would be the claim that every time an agent engages in mindfulness, she executes metacognitive control. I think this is a highly plausible claim, but let me review the two most relevant concerns about it. First, many mindfulness researchers stress that *compassion* is an integral element in mindfulness practice (Goleman & Davidson, 2017). In mindfulness, compassion exercises take two basic forms. In

---

27 However, see Lindhal et al., 2017, for research on psychotic effects of long-term mediation practice.
one kind of practice, agents intentionally establish an attitude of compassion toward individuals or groups. This might involve compassion toward relatives, strangers, or even oneself. In another kind of practice, agents intentionally establish compassion toward their own mental processes. For example, this involves an attitude of compassion toward occurring depressive thoughts or pain experiences, by which you warmly accept and welcome these episodes without any remorse or frustration (Poletti et al., 2021). It is not obvious that we can adequately conceptualize these exercises of compassion in terms of first-order attentional or higher-order attitudinal metacognitive goals. Hence, “Necessary” is too strict: Those mindfulness practices that involve such compassion exercises do not involve metacognitive control as defined by (5)*.

This consideration is important, but we can manage it with minor accommodations. We can easily conceptualize these exercises of compassion by the basic notion of metacognitive goals. For example, we could conceptualize the practice of compassion for other individuals or groups as involving the first-order emotional metacognitive goal of relating with compassion to these individuals (where emotions involve multiple components, such as cognitive, perceptual, attentional, and affective: Scarantino & Sousa, 2021). Further, we could conceptualize the exercise of relating with compassion to one’s own mental processes as involving the second-order attitudinal metacognitive goal of having an attitude of compassion for one’s own mental processes (i.e., a welcoming and accepting attitude toward one’s own mental processes). Implementing these conceptualizations into (5)* would be a minor maneuver. We simply add these compassion-oriented metacognitive goals as disjuncts in the definiens of (5)*.

Second, some readers familiar with non-dual Buddhist traditions, like Zen or Dzogchen, might stress that these traditions do not consider (true) mindfulness as a form of control. Instead, these traditions view true meditative states as absent of any form of agential aspects, as in goals, intentions, and control. I find this consideration relevant, but let me stress that this paper assumes the contemporary scientific approach to mindfulness, in which it is mainstream to view mindfulness as a form of control operation (and, indeed, a form of metacognitive control). The account of mindfulness that this paper offers is an attempt to contribute to this scientific literature, not an attempt to portray ideas of mindfulness in non-dual Buddhist traditions (even though these views are, of course, of clear philosophical interest).

Now, let me sum up these manifold discussions on scope. The positive claim of this paper is that there is one way in which we can give good sense to the idea that the essential and demarcating psychological capacity of mindfulness is a form of metacognitive control. This is
by definition (5)*, which claims that metacognitive control is the implementation of certain goals, namely first-order attentional or second-order attitudinal metacognitive goals. The definition of (5)* (by and large) demarcates mindfulness from most other activities, since many agential activities do not involve metacognitive control. Moreover, the metacognitive control of definition (5)* is also essential to mindfulness in the sense that it is a necessary component. Lastly, the account of (5)* can make sense of the core features of mindfulness (notably, issues relating to explicitness and difference in ultimate goals). This accomplishes more than the preceding accounts reviewed in §3. Consequently, I take it as reasonable to make the conclusive claim that (5)* is the best theoretical account of the metacognitive control of mindfulness that we have. This is a provoking claim, and further discussion might prove it mistaken.

5. Two theses
Now, let me show how definition (5)* motivates two interesting theses on the nature of mindfulness, concerning reductionism and mental action. I think this adds further to the attractiveness of (5)*.

5.1 Reductionism
The question of whether mindfulness is a form of metacognitive control is essentially a question about the ontology of the mind, i.e., the elements of the (human) mind. In general, one can be a primitivist or a reductionist with respect to an element of the mind (e.g., see discussions of self-knowledge: Carruthers, 2011). Being a primitivist about metacognitive control would imply that metacognitive control is a distinct psychological capacity that cannot be reduced to other capacities of the mind. One might claim that this is because metacognitive control involves computational workings or certain neural implementations that are importantly different from those of psychological capacities we already know of. Alternatively, being a reductionist about metacognitive control would imply that we can reduce metacognitive control to other psychological capacities we already know of.

Authors of mindfulness literature do not position themselves explicitly with respect to this ontological issue. Nonetheless, one could interpret some authors as sympathetic to a form of primitivism. For example, Dorjee (2016) proposes that mindfulness is a form of metacognitive self-regulatory capacity. In certain places, Dorjee (p. 4, 5, footnotes 3, 4) seems to argue that although this capacity overlaps with cognitive control and attention control, it cannot be reduced to either of these. Similarly, one might interpret Zawidski (2019) and Leder
and Zawidski (2023) as expressing primitivistic tendencies. Seemingly, they consider mindfulness to be a form of metacognitive control or skill substantially different from other capacities, like ordinary cognitive or attention control involved in perception-motor action. Jankowski and Holas (2014) also appear to claim that the metacognitive control of mindfulness is a form of higher-level processing that is fundamentally distinct from ordinary processing, including cognitive control functioning.

It could be a mistake to interpret these authors as sympathetic to primitivism regarding metacognitive control. Nonetheless, I think there are general reasons for why we should avoid such primitivism. Primitivism is the more expensive position. It claims that the human mind contains a distinct control capacity (i.e., the capacity of metacognitive control) in addition to the control systems of ordinary cognitive control. Further, the primitivist position might imply the idea that even though the human mind contains a latent metacognitive control capacity, agents might never use it or only use it rarely (i.e., only a subset of individuals come to exercise mindfulness or similar practices systematically). This is a strange implication: Why would the human mind contain this latent and distinct control system if it is rarely used? Finally, the primitivist must make sense of why mindfulness practice associates positively with aspects of cognitive control, while claiming that these two capacities of control are distinct and non-reducible.

Definition (5)* offers a way to be reductionist about the metacognitive control of mindfulness. Definition (5)* does not entail the claim that metacognitive control involves unique computational mechanisms or neural implementations that are radically different from those of cognitive control. Instead, (5)* allows us to claim that exercises of metacognitive control involve the same general control capacity as those of cognitive control. The difference is that when agents exercise metacognitive control, they use this control capacity to implement certain metacognitive goals (i.e., first-order attentional or second-order attitudinal metacognitive goals, functioning at the proximal level). This enables a reductive account: Metacognitive control is not a distinct control capacity but a distinct functioning of the general cognitive control capacity. Consequently, we can reduce the psychological capacity of metacognitive control to an agent’s general capacity of cognitive control. Now, some readers might stress that this reductionist account is in tension with the requirements that a satisfying account must enable us to distinguish and explain the difference between metacognitive control and cognitive control (as stressed in §2.1). However, the present reductionist account satisfies this requirement. It states that metacognitive and cognitive control are different because they involve different goals and, hereby, function differently in a very substantial way.
Nevertheless, these capacities are not different primitive capacities or elements of the mind. Notice that this reductionist approach relates back to the discussion of scope in §4.1. Precisely because definition (5)* is a reductive account of the metacognitive control of mindfulness, it is reasonable to expect that other activities could involve this control as well (such as activities of explorative thinking or the recalling of objects). Metacognitive control is not a primitive capacity that is completely unique to mindfulness—nonetheless, at the same time, it is a distinct and relatively rare functioning of our general capacity of cognitive control.

5.2 Mental action

Definition (5)* also brings insights into another issue that concerns seemingly opposing descriptions of mindfulness. Some authors describe mindfulness as a form of nonaction (e.g., Lutz et al., 2019). It is not always clear exactly what this idea of nonaction involves, though it seems to refer to some kind of passive capacity of avoiding acting or reacting to one’s own mental processes. Other authors describe mindfulness as a form of mental action, due to its involvement of control of attention and attitudes (Upton & Brent, 2019; Timalsina, 2022).

Definition (5)* aligns with the second tendency and portrays mindfulness as a mental action. This is clear if we consider the general philosophical idea of such action. The idea of mental action concerns the issue of whether and why some mental operations qualify as actions of an agent (O’Brien & Soteriou, 2009). The idea is surrounded by some controversy. Yet, philosophers seem to agree that the paradigmatic characteristics of mental actions are that (a) these mental operations are intentional under some description, (b) these operations are under the agent’s control, and (c) these operations are covert. The metacognitive control of mindfulness satisfies (a)–(c). When subjects implement first-order attentional or second-order attitudinal metacognitive goals, (a) they do so intentionally, (b) this implementation is a form of control, since it is about relevant goals determining mental processes, and (c) this activity is covert (e.g., agents typically execute the metacognitive control of mindfulness while sitting quietly with closed eyes). Hence, the metacognitive control of mindfulness is a form of mental action. Notice that this is not the same as saying that mindfulness is an ordinary way of acting. As stressed earlier, metacognitive control is not a common ingredient in our everyday undertakings. As a result, when agents enroll in mindfulness protocols, we can say that this involves a kind of substantial “turn” in their agency. Mindfulness protocols turn patients toward

---

28 See also more popular descriptions: https://thedailyguardian.com/meditation-moving-from-action-to-non-action/
29 I draw on Tollerup and Grünbaum (in preparation) in stating these conditions.
forming and selecting goals about their own attentional focus and second-order attitudes and train them to implement these goals. Even though this basic idea of mindfulness as a mental action is not radially new to the literature, definition (5)* offers a new and simple explanation of why this is the case. Mindfulness is a mental action because it involves metacognitive control, i.e., implementation of certain metacognitive goals. This explanation is straightforward and plausible, yet it is not present in current discussions. I think it offers a kind of “demystification” of the nature of mindfulness. Mindfulness is not something esoteric. At its core, it is “simply” the implementation of metacognitive goals.

Nevertheless, notice that the mental action of mindfulness is radically different from the kinds of mental actions that philosophers commonly discuss. One major focus has been on mental actions in relation to epistemic purposes. For example, Hieronmy (2009) argues that agents control belief-like attitudes through the mental actions of exposing themselves to certain evidence or lines of reasoning (see also Proust’s, 2013, account of metacognition in belief-formation). Another focus has been on mental action in practical purposes involving perception-motor control. Numerous philosophers argue that attention control is an integral element in bodily intentional action. They claim that such attention control is a basic form of mental action (Buehler, 2022; Watzl, 2017).

The mental action of mindfulness is importantly different, at least when we consider it in the therapeutic context. When Lucy enrolls in mindfulness to treat her depression, she attempts to attain the relevant proximal metacognitive goals on the background of the ultimate goal of gaining therapeutic effects. This is different from the epistemic purpose of belief-formation, or the attention control integral to bodily action. Lucy’s mental action is ultimately about changing her emotional state or mood. The philosophy of mental action aims to provide a comprehensive account of mental action in its various important functions and basic shapes (O’Brien & Soteriou, 2009). We must include mental action with emotional and mood-oriented purposes in such an account. In fact, this would not be completely new to the philosophical literature. Ancient Greek philosophy contains numerous discussions of mental actions with such purposes (e.g., see Hadot, 1995, on “spiritual exercises”). One might say that the philosophical discussion of mindfulness merely brings back an old philosophical topic: how agents can undertake mental actions to regulate their own emotions and moods.

6. Conclusion
Let me summarize and conclude. §2 provided a review of the scientific literature on mindfulness as a form of metacognitive control. §3 presented the negative claim of this paper,
namely that the preceding accounts of the metacognitive control of mindfulness fail. §4 developed and defended the positive proposal of the paper: that definition (5)* adequately describes the metacognitive control of mindfulness in terms of the implementation of certain metacognitive goals. §5 argued that definition (5)* motivated two novel theses on mindfulness relating to reductionism and mental action. These sections have attempted to give sense to the idea of mindfulness as a form of metacognitive control, present in the interdisciplinary research literature. Yet, the sections have also attempted to provide an account and explanation of mindfulness in terms of contemporary philosophical psychology, in particular through the use of resources from current philosophy of action.

References


Flow and experts

1. Introduction
Philosophers have examined flow in relation to multiple issues, such as the phenomenology of virtuous living (Annas, 2008), or the relation between flow and well-being (Debus, 2015; Haybron, 2020). This paper focuses on the most well-developed and prominent line of philosophical discussion of flow. This is the discussion of flow in relation to the so-called *Automatic view* of expertise (AV) (as labeled by Christensen et al., 2016, p. 41). AV claims that when athletic expertise functions optimally, automatized processes run freely without conscious supervision (Dreyfus & Dreyfus, 1986; Dreyfus, 2002a, 2002b, 2005). One of the key arguments motivating AV stresses the relation between optimal expert performance and flow. Although this argument is not always clearly formulated in the literature, the present paper proposes an explicit outline of it and labels it the *flow argument* (FA). Even though AV is a minority position today, many philosophers and researchers are sympathetic to FA, or at least to central parts of it. The present paper argues that FA is implausible in the light of contemporary psychological research on flow. The paper proceeds in two basic steps. First, it argues that experts undergo psychological states other than flow when they perform optimally. This opposes the first premise of FA. Second, the paper argues that experts in flow execute conscious mental operations, which oppose the second premise of FA. This provides a novel criticism of both FA and AV. Furthermore, it corrects misunderstandings of flow in the broader philosophical and scientific literature.

The paper is structured as follows. §2 outlines FA in an explicit form, consisting of two premises. While §3 targets the first premise, §4 targets the second premise. §5 provides concluding remarks and important topics for further research.

2. The flow argument
The central idea of the Automatic view (AV) is the claim that optimal athletic expert performance is a matter of experts letting automatized processes run without any supervision through conscious control, self-awareness, or task-related thought (Dreyfus & Dreyfus, 1986; Dreyfus, 2002a). Many authors of psychological and philosophical literature have been

---

30 AV acknowledges that conscious supervision might be a part of expert performance but only when such performance functions sub-optimally (Dreyfus & Dreyfus, 1986).
sympathetic to AV (or at least to central aspects of I: See discussions in Christensen et al., 2016). A crucial motivation for AV concerns the position’s ability to account for the psychological state of flow (Christensen, 2019). Although defenders of AV seldom outline their argumentation on flow explicitly, I think the most straightforward way to understand their reasoning is by the following outline. We shall call this the flow argument (FA):

(1) Flow is the characteristic state of optimal expert performance.*
(2) If an agent is in the state of flow, then (a) she does not consciously control her action,
    (b) she is not self-aware about what she does, and (c) she does not have conscious task-related thoughts.
(3) If an expert performs optimally, then she is in a state characterized by claims (a)–(c).

One might be puzzled by the logical structure of the FA. This would concern premise (1) in particular. Premise (1) is underspecified in the above outline (as marked by “*”). This is on purpose. Although it is clear that AV and FA supporters support something like premise (1), it is not obvious exactly what they mean by such a claim. Hence, premise (1) remains underspecified in the initial outline of the FA. The following §3 specifies what AV and FA supporters could mean by premise (1), on a more precise level, proposing that one should specify premise (1) as a conditional. For now, let me just highlight that AV and FA supporters clearly commit to the basic idea of premise (1) (this is also suggested by other authors, e.g., Christensen, 2019; Fridland, 2021). For example, Dreyfus (2007b, p. 373) writes:

Experts experience periods of performance, variously called “flow”, “in the groove” and “in the zone”, when everything becomes easier, confidence rises, time slows down, and the mind, which usually monitors performance, is quieted. Yet performance is at its peak.


Turning to premise (2), this involves three negative claims about the state of flow. Claim (a) states that when an agent is in flow she does not consciously control her action. Control of action is obviously a complex matter. Defenders of FA and AV seem to claim that agents in flow do not consciously supervise their actions by guiding, evaluating, or consciously
implementing intentions or goals. When in flow, perception-motor processes unfold automatically without any need for conscious initiation or ongoing regulation.

Claim (b) is also unclear in several respects. Self-awareness is a major philosophical topic, and philosophers distinguish between multiple kinds of such awareness. As other philosophers indicate, it is not clear what kind of self-awareness FA and AV exclude from flow states (Dow, 2017). I will understand claim (b) as excluding any kind of ongoing conscious monitoring of what one is doing and how successful one is in doing it.

Claim (c) states that flow does not involve any task-related thoughts. The qualification of “task-related” should be stressed. Claim (c) does allow for random thoughts to “pop up” under performance (e.g., an athlete might suddenly think about buying groceries while performing). What (c) does not allow is athletes thinking about their ongoing performance while being in flow. This would especially concern evaluative thoughts about the success of one’s own performance and strategic thoughts about how to act.

Claims (a)–(c) are obviously interrelated. One might think they imply each other in some way. I outline them separately for the sake of clarity. AV and FA supporters often endorse these claims in conjunction with each other. Such a conjunction might be ordered under the general slogan of AV, namely that the optimal performance of expert skill is mindless (Dreyfus & Dreyfus, 1986). For example, Dreyfus (2007a) thinks that the content of flow experience is solely made up of attracting and repelling forces. There is no conscious guiding, planning, self-awareness, or thought present in this experience. Dreyfus writes: “when one is in flow as athletes say—there is no experience of this body as mine but simply the experience of ongoing coping … in the experience itself no ‘I’ was present” (2007a, p. 356). This view is echoed in other passages where Dreyfus discusses flow, claiming that: “Athletes in such situations say they are playing out of their heads … In general, when one is totally absorbed in one’s activity, one ceases to be a subject” (2007b, p. 373). Brownstein follows Dreyfus’ phenomenological account of flow by claiming that when in flow an agent “is not thinking about herself, or even thinking about anything at all, while she acts” (2014, p. 546). He continues by describing flow as follows:

... expert action unfolds in the absence of self-reflective thought or conscious self-awareness… The phenomenon of flow—of being carried forward unreflectively in the performance of a difficult action—is hard to understand (p. 546; see also, p. 559)
The present paper aims to provide a novel critique of FA, and hereby AV. Criticism of AV is not new. In fact, one might say that the core thread of contemporary philosophy of skill and expertise is the criticism of AV. Yet, the preceding criticisms of AV have not devoted themselves to scrutinizing FA specifically. Instead, philosophers have argued that AV is implausible with respect to general aspects of cognitive control and conscious processing in expertise (Christensen et al., 2016; Montero, 2016; Fridland, 2017). This lack of exclusive focus on FA is unfortunate for two reasons. First, FA is a main argument motivating AV (Christensen, 2019). Analyzing the exact workings of FA is important for evaluating AV. Second, though the general view of AV might not have persuaded many contemporary philosophers of skill and expertise, AV has been very successful in advocating FA. The result seems to be that philosophers who are otherwise not proponents of AV appear sympathetic to FA’s descriptions of flow (e.g., Velleman, 2008; Railton, 2009; Annas, 2008). Many psychologists and popular writings are also sympathetic to FA (e.g., Duckworth, 2016, pp. 129–33). Hence, showing that FA is implausible would correct misunderstandings of flow in the wider philosophical and scientific community.

Lastly, I should mention that this paper assumes that FA and AV supporters refer to the psychological construct of flow in their argumentation (and not some idiosyncratic notion). This assumption is reasonable, both in a descriptive and normative sense. Descriptively, FA and AV supporters refer continuously to the psychological work of Csikszentmihalyi (1990) to support their argument. Hereby, they seem to refer to the psychological construct of flow. More normatively, it is reasonable to demand that these supporters refer to this psychological construct. Performance and sports psychology has aimed to develop a valid construct of flow with associated scales for measurement. If FA is to tell us anything about the nature of expert performance in terms of flow, it should operate with a validated scientific notion of flow and not some ill-defined or idiosyncratic notion that does not correspond to the relevant psychological research. Otherwise, FA and AV supporters should explicitly explain their choice of a divergent notion.

3. Flow and optimal performance

Let us begin by considering premise (1), stating that flow is the characteristic experience of optimal expert performance. Although I shall specify (1) into two possible conditionals, I do

31 Notice that psychologists measure flow in two different ways: While dispositional flow scales measure individuals’ disposition to get in flow, state flow scales measure the level of flow in a given situation (Jackson & Eklund, 2002).
not interpret (1) as stating a strict logical relation between flow and expert performance (as in claiming some relation of logical sufficiency or necessity between the two). This interpretation would be inappropriate. FA and AV supporters aim to provide a descriptive phenomenological and psychological account of athletic expert performance that sketches its general character (Dreyfus, 2002b, 2007a, 2007b). Hence, one should understand the following conditionals as expressing underlying statistical claims.

### 3.1 Flow as exclusive to experts

One way to specify premise (1) would be to interpret it as the claim that flow is exclusive to experts that perform optimally. Following this, we could specify premise (1) as the following conditional: If an agent is in flow, then she is an expert performing optimally. This yields a prediction: Every time we find an agent in flow, this agent will be an expert performing optimally. Supporters of FA and AV make claims that indicate commitment to this conditional and prediction (Dreyfus, 2002a, 2002b, 2007a; see also Brownstein, 2014, p. 548). Numerous other philosophers do as well (Annas, 2006; Garfield & Priest, 2020).

Would premise (1) be plausible if we specify it in this way? No. Two considerations suggest that this conditional is deeply problematic.

First, one obvious problem is that you cannot reach the conclusion in (3) with this interpretation of premise (1) (assuming that we understand FA as a transitive argument aiming for validity). Paired with premise (2), this conditional does not imply the claim of (3): Premises (1) and (2) would be fully compatible with experts being in other states than that of flow, and these states are not characterized by (a)–(c). Hereby, the present interpretation of premise (1) carries an inferential problem: It does not involve the right kind of quantification to draw the desired conclusion in (3). This is a point worth stressing, since AV and FA supporters seem to overlook it (such as in Dreyfus, 2002a, 2002b, 2005).

Second, the conditional is also empirically implausible. I should mention that this taps into a general problem in the psychological research on how to understand the relation between flow and skill level generally. It traces back to unclear aspects of Czikszentmihalyi’s (1990) theorizing. His writings formulate flow as both a kind of rare peak experience of experts and as a much more mundane state open to everyone in any kind of activity. Nonetheless, considering contemporary empirical research, flow does not seem exclusive to experts. For example, studies in education and learning suggest that flow facilitates students’ skill acquisition (e.g., Rossin et al., 2009). Other studies highlight the role of flow in creative thinking, basic physical exercise, consumer experience, or other ordinary or everyday activities.
among individuals who are not experts in the respective domains (unless we operate with a very thin notion of expertise. See Primus & Sonnenburg, 2018; Carter et al., 2013; Asakawa, 2004; Lavoie et al., 2021). Such research shows that nonexperts undergo flow states.32

3.2 Optimal performance and flow

Alternatively, we can interpret premise (1) as stating the reversed conditional: If an expert performs optimally, then she is in the state of flow. This yields another prediction: Whenever we find an expert performing optimally, she will be in flow. AV supporters seem to support this claim (e.g., Dreyfus, 2002b, 2007a). Other philosophers (e.g., Papineau, 2013), neuroscientists (van der Linden et al., 2021), and psychologists (Duckworth, 2016, pp. 129–33) do as well. This specification of premise (1) also avoids the inferential problem above. It is the kind of conditional you need to make FA into a valid transitive argument, enabling the inference of (3).

However, psychological research also suggests this specification of premise (1) to be empirically implausible. To narrow our focus, let us focus on the seminal research by Swann et al. (2016, 2017a, 2017b). Through multiple empirical investigations with different athletic populations, Swann and colleagues found that athletes and experts undergo two different paradigmatic states when they perform optimally. Athletes describe both of these states as states of “being in the zone,” i.e., as states of being deeply present in their performance. As for the first state, athletes reported that optimal performance was a matter of “letting it happen.” Here, “being in the zone” was a matter of not breaking the fluency of optimal performance (Swann et al., 2016). Swann and colleagues argued that this was in essence the state of flow.

As for the second state, athletes reported another kind of “being in the zone,” where they described performance as a matter of “making it happen.” This was a state in which optimal performance had to be established in a more effortful way. Swann and colleagues labeled this state clutch. Under clutch, athletes reported that they consciously deliberated about what to do, they exerted explicit and intense effort in reaching their goal, and they had a clear awareness of the pressure of the performance situation and their own position within it. The research by Swann et al. (2016, 2017b) suggests that clutch states are equally characteristic for optimal expert performance as that of flow. Moreover, Swann et al. (2017a) proposed that flow and clutch states often interact under optimal performance depending upon the conditions of the

---

32 An alternative interpretation would be that premise (1) makes the claim that experts have the highest degree of flow (acknowledging that non-experts can have some lower degree). Nevertheless, this interpretation still faces the inferential problem above, and it is neither obvious that it is empirically warranted.
external environment, the success of the performance, psychological traits, and the cognitive and metacognitive strategies of the performer (see also Brick et al., 2019).33

Following on from this, flow does not exhaust the psychological and phenomenological nature of optimal athletic expert performance. The fact that clutch is equally characteristic of optimal expert performance shows that the present specification of premise (1) is false. An expert performing optimally might also, and with high probability, be in clutch. Since clutch and flow, by their respective psychological features, are often exclusive of each other, this implies that the expert is not in flow. Moreover, it is clearly hard, if not impossible, to make sense of clutch through the lens of AV. Recall that AV claims that optimal expert performance is a matter of letting automatized processes run without conscious supervision. However, clutch states involve conscious supervision: Athletes consciously plan, reflect, motivate, initiate, and evaluate themselves and their own situation while being in clutch (Swann et al., 2017). Hereby, the AV is unable to account for the mental complexity and conscious aspects of clutch states, since it describes optimal expert performance as a state of unreflective and totally automatized activity.

With the above reasoning in mind, the broader lesson is that a philosophical theory of expertise, which aims to provide an adequate account of the psychology of optimal athletic expert performance, must account for both flow and clutch as well as their interaction. This might surprise some philosophers. Authors often understand the expression of “being in the zone” as synonymous with flow (e.g., Papineau, 2013).34

4. The psychological state of flow
Consider premise (2) of the FA. It states the conditional that if an agent is in a state of flow, then (a) she does not consciously control her action, (b) she is not self-aware of what she is doing, and (c) she does not have conscious task-related thoughts. The following sections develop three lines of objections to this premise.

33 Notice that Mylopoulos and Pacherie (2021) use the exact same expressions of “letting it happen” and “making it happen” to depict two basic modes of expert performance. It is not clear whether they refer to the constructs of flow and clutch.

34 However, notice that other philosophers have already stressed that flow does not exhaust the psychological and phenomenological nature of optimal athletic expert performance. For example, Shepherd (2021) stresses that current psychological studies only suggest a very modest correlation between flow states and objective performance (see Kennedy et al., 2014). Other philosophers have made related points (Christensen et al., 2016; Hoffding, 2019; Montero, 2016). Nevertheless, their argumentation does not involve the psychological construct of “clutch.” This construct generally seems underappreciated in contemporary philosophy of skill and expertise (although see also Toner & Moran, 2021).
4.1 Multi-dimensional construct

Flow is most commonly conceptualized as a multi-dimensional construct in psychology (Jackson & Eklund, 2002). Performance and sports psychologists conceptualize the flow construct as involving three so-called flow conditions and six flow characteristics (in the following we shall only focus on a subset of the characteristics).

The nine dimensions of flow

*Flow conditions*: 
1. Challenge-skill balance
2. Clear goals
3. Unambiguous feedback

*Flow characteristics*: 
4. Sense of control
5. Concentration
6. Action and awareness merging
7. Loss of self-consciousness
8. Autotelic experience
9. Time transformation

Measuring flow states involves the measurement of each of the nine dimensions of flow. Scores of each dimension are then aggregated to compute a global flow score. Consequently, two individuals can score high in their global flow score, yet have very little overlap in their scores on separate flow dimensions. Flow is, hereby, a heterogeneous and global kind of state. Heterogeneous in the sense that it can have multiple forms and structures. Global in the sense that it is composed of different dimensions. If we accept that flow is a multi-dimensional construct, it is clear that referring to flow means referring to the three conditions, the six characteristics, and their many ways of combining.

This multi-dimensional nature of flow is important for assessing premise (2) in the FA. If this premise and the claims of (a)–(c) are to be accepted, they must adequately cohere with these flow dimensions. Let us begin by stressing that some flow dimensions might, at first sight, support premise (2). These would be the flow dimensions of “action-awareness merging” and “loss of self-consciousness.” Indeed, when philosophers have argued in favor of premise

---

35 For alternative conceptualization, see Marty-Dugas and Smilek (2019) or Lavoie et al. (2022).
(2), they have drawn on anecdotal reports of athletic experts, in which athletes describe their optimal performance in ways that align with these two dimensions (Brownstein, 2014; Dreyfus, 2005).

Yet, it is not obvious that these two dimensions univocally support premise (2). This is clear when we consider how they are operationalized and measured by scales such as the widely used Flow State Scale by Jackson and Marsh (1996). Take the dimension of “action-awareness merging.” This dimension is measured by rating one’s agreement with statements like “I performed automatically” or “I did things spontaneously and automatically without having to think.” The question becomes: Does agreement with these statements univocally support premise (2)?

I think this depends upon one’s background view concerning the dynamics between conscious control and automaticity in action control. Philosophers and empirical researchers generally view action control as multileveled. This means that actions, such as those involved in athletic performance, involve interactions between higher-level psychological processes (involving intentions, beliefs, thoughts, perceptions, etc.) and lower-level bodily and physiological processes (muscle dynamics, fine-grained correction of movement, etc.) (Pacherie, 2008). Take the example of Wu’s (2011, 2016, 2020) account of the relation between automaticity and control. According to this theory, a single perception-motor activity can be automatic at some level (e.g., the fine-grained motor or eye movements) while at the same time being consciously controlled at another level (e.g., through intentions, attention control, and monitoring of success). Similarly, consider Christensen et al.’s (2016) framework of skilled action. This framework involves three general control levels: strategic control (concerning control in terms of the overall goal, e.g., controlling a car to reach a destination), situational control (control in relation to the immediate situation, e.g., navigating the immediate traffic), and implementation control (control of actions that achieve situational control, e.g., changing gears, steering). These authors argue that cognitive control is involved in strategic and situational control, drawing on multiple empirical studies. Other philosophers express similar multilevel views of skilled agency and expertise (e.g., Shepherd, 2019; Fridland, 2017, 2021; Pacherie & Mylopolous, 2021). Empirical researchers do as well (Bläsing et al., 2009).

The important thing is that the flow dimension of “action-awareness merging” does not by itself imply that the automaticity of flow is so radical that it excludes conscious supervision at any processing level. Agreement with statements like those above is fully compatible with thinking that flow involves conscious supervision at higher levels. This would indeed be a reasonable idea given the widespread multilevel theories of action control. Under these
theories, we should expect that the “action-awareness merging” dimension concerns automaticity at lower and intermediate levels (many flow researchers actually follow this view: Csikszentmihalyi & Nakamura, 2010). Defenders of the FA have to provide a strong story about why the automaticity of the “action-awareness merging” is so radical that it excludes conscious control at any level.

Then, recall the dimension of “loss of self-consciousness.” Again, it is not obvious that this dimension supports premise (2) when we consider its operationalization and measurement. The dimension is measured by subjects rating their agreement with statements like “I was not worried about what others may have been thinking of me.” This obviously refers to a very specific kind of self-consciousness, i.e., a social kind of self-consciousness (as explicitly recognized in Nakamura & Csikszentmihalyi, 2014, p. 240). Most philosophers believe that human self-consciousness is not exhausted by social or reflective forms. They claim that the human mind also involves more pre-reflective forms of self-consciousness and self-awareness that are implicit in bodily action. The flow dimension of “loss of self-consciousness” does not imply a loss of such pre-reflective forms (Dow, 2017). Actually, some psychologists argue that these forms of self-awareness are integral to flow (Toner et al., 2016, p. 306).

This critique can be bolstered by considering the relationship between mindfulness and flow. It is generally underappreciated in philosophical discussions that mindfulness practice (as one of the few interventions) enhances the flow scores of athletes (Kee & Wang, 2008; Cathcart et al., 2014; see review by Goddard et al., 2021). The important thing is that psychologists and cognitive scientists understand mindfulness practice as increasing individuals’ ongoing self-awareness of their own mental processes, involving their attention allocation and emotions. Hence, one very natural explanation of the positive relation between mindfulness and flow would be that this self-awareness aids attention control and ensures full concentration, hereby fostering flow (Brick et al., 2019). Again, FA supporters have to tell us why proposals like this are mistaken.

Things bring us to another flow dimension, namely that of “concentration.” This dimension refers to flow states being states of heightened and sustained concentration. Although contemporary psychologists disagree on how to conceptualize flow in certain respects, they do agree that high concentration is a, if not the, key dimension of flow (Marty-Dugas & Smilek, 2019; Lavoie et al., 2022). Drawing once again on mainstream background theories, it is clear that this dimension is most likely troublesome for premise (2) in the FA.

Here is one way in which this could be the case. One might think that heightened and sustained concentration of flow demands person-level attention control and that such control
involves conscious control and self-awareness. This view on concentration and attention control is widespread in the philosophical literature (see also Toner & Moran, 2021; Harris et al., 2017). Many contemporary philosophers think that attention control involves agents actively and consciously structuring and sustaining their attention in accordance with their relevant intentions and goals. Various philosophical theories state this ability as a core element in agency and skilled action (Watzl, 2017; Wu, 2011, 2016; Buehler, 2021). According to such a view, an instance of behavior can only qualify as being an intentional action if it is mediated or enabled by attention that is constrained by the agent’s relevant intentions or goals. This is also relevant for understanding skilled action. Although philosophers commonly recognize that skilled action does indeed involve automatized and embodied processes, it also depends upon sustained attention to the relevant task and its many aspects. Such attention is constrained by an agent’s intentions and goals. This involves conscious control in the form of continuous top-down constraining of one’s own activity (Bermúdez, 2017). According to this type of philosophical theory, it would not make sense to propose that the heightened and sustained concentration of flow did not depend upon on some form of conscious control in the form of top-down attention control. Drawing on this type of theory, the flow dimension of “concentration” would likely be in conflict with sub-claim (a) in premise (2) (see also Moran et al., 2019).

Moreover, multiple philosophers have presented strong arguments for the idea that such person-level attention control always involves some kind of self-awareness (Watzl, 2017; Metzinger, 2013). The idea is that agents can only control their attention, such that it aligns with their intentions or goals, by having a fluent awareness of their own attentional orientation. This self-awareness is not explicit in the sense that agents constantly and consciously make judgements about their own attentional orientation. It is instead more subtle and implicit. Nevertheless, this plausible idea opposes sub-claim (b) of premise (2) by suggesting that flow does involve self-awareness—namely, self-awareness of one’s own attentional allocation.

To conclude, there is a lot of room for maneuver to make sense of flow dimensions in ways that are troublesome for premise (2) of FA. Drawing on independent and mainstream philosophical and psychological background theories on the multileveled nature of action control, the multiplicity of self-consciousness and self-awareness, and the nature of attention control, the flow dimensions of “action-awareness merging” and “loss of self-consciousness” do not univocally support premise (2). These conclusions stress an important and general point: Flow dimensions do not in isolation favor certain philosophical theories of expertise. One must couple these dimensions with relevant background theories to draw these wider conclusions.
AV and FA supporters have neglected this point. They have portrayed the phenomenon of flow as independently supporting their own view of expertise without any need for making assumptions on background views.

Notice that we can back up this theoretically oriented criticism of premise (2) with more empirical considerations. Consider, for example, Swann et al.’s (2012) systematic review of sports psychological flow studies. This review showed three things of interest.

First, Swann et al. (2012) found that the dimension of “loss of self-consciousness” was among the weakest dimensions in flow states among athletes (fewer than 30% of athletes reported this dimension as increased under flow). In fact, increase in this dimension is so rare that some psychologists question whether it should be excluded from the flow construct altogether (see Swann et al., 2018, for useful discussion). This is clearly bad news for FA supporters. They have often argued that a lack of self-awareness is a characteristic, if not a universal aspect, of flow.

Second, Swann et al. (2012) found that the most common dimension that athletes reported from their flow experience was high concentration on the present task (80% reported this to be a part of their flow experience). If one is sympathetic to the idea that sustained concentration depends upon a kind of person-level and conscious attention control that involves ongoing self-awareness, this observation strengthens the pressure on premise (2).

Third, Swann et al. (2012) found that the second most common feature of flow was the occurrence of positive thoughts and emotions (75% of athletes reported this to be a part of their flow experience). Although Swann et al. do not explicitly specify what they mean by a “thought,” it seems reasonable to think that these positive thoughts involve task-related conscious mental states with some conceptual or directly linguistic content, i.e., thoughts like “you are doing good” or “things are going according to plan” (what should they otherwise be positive thoughts about if not the relevant activity?). This seems to put pressure on sub-claim (b): These thoughts constitute a kind of self-awareness about one’s own success. They also put pressure on sub-claim (c) (the claim that flow states do not involve any task-related thoughts). The regular presence of positive thoughts in flow show that task-related thoughts occur while being in flow.

Note that this consideration of task-related thoughts relates to the phenomenon of self-talk in expertise. Self-talk is a kind of self-directed inner speech that many elite athletes draw on to support their performance. Such self-talk can be both motivational (in which its purpose is to support and sustain goal-directed behavior) and instructional in nature (in which its purpose is to strategically guide goal-directed behavior). Motivational self-talk involves
evaluative thoughts about one’s own performance (e.g., “Keep on going; you’re doing well!”),
while instructional self-talk can involve more strategic thoughts (e.g., “Hold your knee at that angle.”) (Bellomo et al., 2020). The role of motivational self-talk is not commonly discussed in the philosophy of skill and expertise. However, contemporary performance and sports psychologists think that such self-talk can be crucial to optimal performance states, involving flow, in many different athletic domains (Brick et al., 2019). This puts considerable pressure upon claim (c) of premise (2). Again, it stresses that there is general and empirical support for proposing that task-related thoughts often occur under flow states, namely in the form of self-talk.

3.2 Absorption

How could the FA supporter respond to the above objections? The most promising way might be to argue that these objections totally neglect the fact that optimal athletic expert performance is characterized by absorption. This casts new light on the psychological nature of flow (Dreyfus, 2002b, 2007a; Brownstein, 2014). The FA supporter could reason that states of absorption are exactly states in which agents are fully immersed in their activity without conscious supervision, resulting in some trance-like state. This would mean that states of absorption are generally characterized by premise (2) and claims (a)–(c); i.e., they are states without conscious control, self-awareness, or task-related thoughts. One might find similar descriptions among psychologists and neuroscientists that describe flow as a state of absorption (e.g., Lavoie et al., 2022; van der Linden et al., 2021). Even philosophers opposing AV make such descriptions (e.g., Fridland, 2021; Shepherd, 2021; Toner et al., 2016).

Nonetheless, I think there are good reasons for being skeptical toward this idea of flow as a state of absorption. Initially, we might ask the FA supporter: What does the idea that flow occurs while in a state of absorption more precisely concern? What does “absorption” refer to in this idea? Psychologists have developed the notion of “absorption” into a psychological construct involving a distinct conceptualization and scales for measurement (notably, Tellegen & Atkinson, 1974; Tellegen, 1981). The FA supporter should draw on this scientific construct. Hence, we can specify the matter in more precise terms: Are the constructs of flow and the construct absorption as deeply related as FA supporters, and many other authors, suggest them to be?

One recent psychological study can aid us in answering this question. In their psychometric validation study, Marty-Dugas and Smilek (2019) investigated the relationship between scores on flow scales and scores on standard absorption scales. The authors found a
significant, but only modest, positive correlation between these scores (correlation between the
two different flow measurements and scores on the Tellegen Absorption Scale: $r = .191$, $r =
.224$, both $p < 0.01$). Following common practice in psychology, if flow typically occurs while
in a state of absorption, we should expect this correlation to be considerably stronger. We can
explicate this difference between flow and absorption further by considering their respective
relations to so-called everyday inattention (an individual’s tendency to forget stuff, mind-
wanter, etc.). In Marty-Douglas and Smilek (2019), measurements of absorption correlated
significantly and positively with everyday inattention (correlations in the range of .365 to .45,
$p < 0.01$, for different measurements). This is in sharp contrast to the correlational patterns of
flow. They correlated significantly and \textit{negatively} with everyday inattention (correlations in
the range of -.286 to -.376, $p < 0.01$, for different measurements). These results led Marty-
these authors to suggest that the reason why flow and absorption might initially seem related is that
they are both states of intense, exclusive focus on some task or activity. Nonetheless, these
types of focus seem to have very different characters, given that they have such different
associations with everyday inattention. It is beyond the present scope to explore this difference
further, but it is indeed a topic worth future work in philosophy and psychology. My own
speculation would be that absorption involves a more stimulus-based focus (where external
objects or spontaneous mental states draw individuals’ focus toward them), while flow involves
a more goal-based focused (where individuals’ goals structure focus—see also §3.3 below).36

These considerations stress that flow and absorption are distinct constructs. Although
they are (modestly) related by their common involvement of intense focus, the attentional
nature of this focus seems importantly different. This is critical. Claiming flow to be a state of
absorption is extremely common. This description has given unwarranted support to AV and
FA by implicitly suggesting flow to be a trance-like state. The relationship between flow and
absorption seems much more complex, and authors should generally be careful in stressing this
relationship.

Now, the FA supporter might simply respond that her notion of absorption does not
correspond to the psychological construct of absorption, and therefore the above argument does
not apply. This reply is open; however, it does not remove the burden of the FA supporter to
formulate the concept of absorption that she finds characteristic of flow. Notably, she must

\footnotesize{36 Note that Marty-Dugas and Smilek (2019) measured flow as a one-dimensional construct, although many of
the items in their scale resemble those of traditional flow scales that measure flow as the nine-dimensional
construct outlined previously. Importantly, they do \textit{not} propose to reconceptualize flow in terms of absorption
(as Shepherd, 2022, part 2.1, seems to suggest). Marty-Dugas and Smilek (2019) propose to reconceptualize
flow as \textit{deep, effortless concentration}.}
distinguish this concept of absorption from the standard psychological construct, and she must explicate why her notion of absorption is not inappropriately idiosyncratic or obscure. More generally, her concept of absorption must also aid her in tackling the criticism of premise (2) in §3.1 above, concerning the multidimensional nature of the current flow construct. This is a considerable task for future discussion. For now, we can conclude that the FA supporter cannot draw on the psychological construct of absorption to support premise (2).

3.3 Mental toughness

I now turn to the final critique of premise (2) of FA. I do not view this critique as a direct refutation of premise (2) but rather as a formulation of a novel and important challenge.

The challenge concerns the relationship between flow and so-called mental toughness. Psychologists generally describe mental toughness as the capacity to maintain goal-driven behavior despite hindrances (Connaughton et al., 2008, p. 193). One prominent approach is to conceptualize mental toughness as a four-dimensional construct, involving the four Cs. This refers to a certain psychological mindset. The mental toughness of an agent is a matter of how much the agent acts and thinks with the view that she can influence the relevant situation (control), trusts her own capabilities to succeed in the relevant task (confidence), views challenges as opportunities to grow (challenge), and is dedicated to succeed in the task (commitment) (Clough et al., 2002). Hence, mental toughness is different from physical toughness. We might say that physical toughness concerns the capacity of an agent’s body to perform under physical pressure (e.g., lifting weights, processing lactic acid, level of respiratory intensity). Importantly, such physical toughness and mental toughness interact in athletic performance. One agent, A, might outperform another agent, B, in some physical task, even if B’s physical capacity is superior to A’s. This is because agent A has a higher degree of mental toughness, hereby maintaining her goal-driven behavior to a greater degree (making more use of her physical capacities than B).

Multiple studies show a strong positive relation between flow and mental toughness. Meggs et al. (2019) found that measures of mental toughness correlated positively with disposition to get in flow among ironmen and standard distance triathletes ($r = .67$ to $.81$, $p < 0.05$). Jackman et al. (2017) found a positive correlation between measures of mental toughness and dispositions for flow in university athletes ($r = .5$, $p < 0.01$). Further, Crust and Swann (2013) found a positive correlation between measures of mental toughness and global flow score ($r = .65$, $p < 0.01$). With such results, there is the widespread idea in the psychological
literature that the psychological capacity of mental toughness contributes positively to making agents more prone to flow states.\textsuperscript{37}

Obviously, these are merely correlations, and one might stress that they do not entail a causal relationship between mental toughness and flow (the correlation might be confounded by other phenomena). However, it does seem highly plausible that a causal relationship exists—at least, this is the standard view in psychology (see the studies cited above). Now, there are two basic ways in which we could account for such a causal relationship between mental toughness and flow. This concerns their temporal relation. First, one could understand the relationship as \textit{diachronic} in the following way: The capacity of mental toughness contributes to flow states prior to the flow states themselves. For example, one might think that the mindset of mental toughness plays an important role when athletes set their goals before they perform (e.g., when an athlete says to herself, “This time you are going to break your personal record” before the performance itself). Such goal setting would be a purely preparatory process before performance itself. Second, one could understand the relationship as \textit{synchronic} in the following way: The capacity of mental toughness contributes to flow states under the flow states themselves. For example, one might think that the capacity of mental toughness continuously counteracts rumination, worry, and other mental processes that can occur under flow states—processes that could potentially disrupt the flow states if not managed sufficiently. Such an account stresses that the mindset of mental toughness is active while athletes are in flow and that the mindset supports the continuation of the flow state.

Which account should we favor—or should we commit to both? Initially, let me highlight that the diachronic account coheres perfectly well with the FA, and the AV more broadly. The FA and the AV make claims solely about what happens \textit{under} expert performance itself, hence flow states themselves. They make no claims about the psychological dynamics leading up to performance or flow states. Some psychologists seem to think that mindsets, like mental toughness, do make these diachronic contributions (Jones et al., 2007; see also Gollwitzer, Smith et al. (2020) examined the relation between the construct of “grit” and flow (for philosophical discussion of grit, see Morton & Paul, 2019). “Grit” is usually conceptualized as “perseverance and passion for long-term goals” (Duckworth et al., 2007, p. 1087), which seems to make this construct overlap with that of mental toughness. Smith et al. (2020) found two interesting correlations. First, they found a negative relation between grit and different kinds of inattention (correlations ranging from around -.20 to -.40, p < .01). Recall that absorption, on the other hand, was positively correlated with inattention. Second, they found a positive relation between grit and deep, effortless concentration (correlations ranging from around .20 to .38, p < .01). They concluded that the results support the idea that flow is strongly and positively related to capacities for long-term goal attainment.

\textsuperscript{37} Smith et al. (2020) examined the relation between the construct of “grit” and flow (for philosophical discussion of grit, see Morton & Paul, 2019). “Grit” is usually conceptualized as “perseverance and passion for long-term goals” (Duckworth et al., 2007, p. 1087), which seems to make this construct overlap with that of mental toughness. Smith et al. (2020) found two interesting correlations. First, they found a negative relation between grit and different kinds of inattention (correlations ranging from around -.20 to -.40, p < .01). Recall that absorption, on the other hand, was positively correlated with inattention. Second, they found a positive relation between grit and deep, effortless concentration (correlations ranging from around .20 to .38, p < .01). They concluded that the results support the idea that flow is strongly and positively related to capacities for long-term goal attainment.
However, research does not suggest that mindsets like mental toughness *solely* make diachronic contributions. Research on mental toughness most commonly highlights the role that mental toughness plays under performance itself, i.e., its synchronic contribution (Gucciardi et al., 2016; Crust & Clough, 2011). The natural environment of optimal expert performance is that of intense competition, involving many hindrances (Christensen et al., 2016). Studies highlight how the capacity of mental toughness is crucial in managing these hindrances as they present themselves. When athletes are in flow and face these hindrances, the capacity of mental toughness counters that the hindrances do not disrupt the concentration, motivation, and confidence crucial for continuing the flow state. For example, a tennis player in flow might make a bad shot or lose a point. If she does not cope with this negative event sufficiently, it might disrupt her state of flow. However, being mentally tough will allow her to continue the flow state by ensuring focus, confidence, and motivation despite occurring hindrances.

In sum, the above reasoning has argued for three claims:

(i) The capacity of mental toughness is a mindset that enables agents to maintain goal-driven behavior despite hindrances.

(ii) There is a causal relation between the capacity of mental toughness and flow states, such that mental toughness contributes positively to flow states.

(iii) A significant part of the contribution of mental toughness to flow is synchronic.

All three claims are highly plausible. Let me show why they present a challenge. The FA supporter must explain why the synchronic contribution of mental toughness to flow is not in tension with the FA, or the AV more broadly. This is a difficult task, for at least two reasons.

First, there are good reasons to believe that the synchronic contribution of mental toughness to flow *does* involve conscious processing, opposing premise (2) in FA. Multiple studies have investigated performance situations involving mental toughness (most often, through in-depth interviews with athletes). These studies stress how mental toughness involves various workings. For example, mentally tough athletes actively interpret feedback in the environment in ways that motivate their performance; they intentionally focus upon the moment and acute tasks instead of worrying about distant outcomes or ruminating about past events; and they actively approach their own emotions and affects in ways that are fruitful for confidence, e.g., by interpreting their high arousal as readiness rather than nervousness (e.g., Gucciardi et al., 2008; Jones et al., 2007; Jaeschke et al., 2016). These workings would counter
premise (2) of FA. Actively interpreting feedback, intentionally limiting one’s focus, or approaching one’s own emotions seemingly involve aspects of conscious control, self-awareness, and task-related thoughts. At the very least, the FA supporter has to tell us why this is not the case.

Second, even if one assumes the FA supporter could come up with a reasonable response to this challenge (which I doubt), the supporter faces another problem. This concerns AV as a broader position. Recall that the AV, as a general theory, claims that optimal expert performance and flow is a matter of letting automatized processes run without conscious supervision. AV supporters describe these automatized processes as embodied processes (Dreyfus, 2002a, 2002b, 2007a, 2007b). Yet, if one recognizes that mindsets like that of mental toughness make a synchronic contribution to performance and flow states, this is seemingly in tension with the central claim of AV. Mental toughness is a psychological capacity; it is a mindset. If such a capacity is active under performance and flow, it is inadequate to claim that expert performance is merely a matter of letting such automatized processes run without conscious supervision. Given the role of mental toughness, expert performance would also be a matter of having and executing certain psychological capacities. It stresses a psychological dimension of optimal expert performance, which AV seems to neglect—or, at least, does not describe adequately.

4. Concluding remarks
This paper began with outlining the flow argument (FA), which plays a central role in motivating the Automatic view of expert skill (AV). I argued that premise (1) of FA makes unwarranted claims about the nature of optimal expert performance (mostly due to neglecting the state of clutch). Further, I argued that premise (2) is implausible given the multidimensional nature of the flow construct and various relations to other constructs, such as absorption and mental toughness.

This paper has made the claim that flow, as a type of state, is not portrayed adequately in FA. My argument allows that some tokens of flow might have the trance-like characteristics expressed in FA (in accordance with the expert anecdotes in Brownstein, 2014, and Dreyfus, 2007a). However, the psychological research suggests such instances of flow to be rare. Consequently, just like Dreyfus (2005, 2007a) claimed that a “myth of the mental” dominated the philosophy of skill and expertise, this paper has argued that a “myth of flow” has dominated the philosophy and science of flow. This myth is prevalent precisely due to an overreliance on
specific instances of flow expressed in anecdotal reports by experts rather than a careful examination of the relevant psychological literature.

Finally, let me stress that the present paper does not exhaust the ways in which the current psychological literature enables novel arguments against FA and AV. For example, I have not discussed the increasingly influential idea that certain kinds of metacognition are important for athletic performance and flow states. Numerous psychologists argue that athletes use metacognition prior to and during performance. Examples concern strategic control of attention, or mental techniques in managing pain or fatigue (Moran et al., 2019; Brick et al., 2015; MacIntyre et al., 2015; Brick et al., 2020; Brick et al., 2019). This is interesting. Standard objections to the AV normally stress that conscious supervision is important in perception-motor processes between the athlete and her environment. The present case of metacognition makes another claim: It stresses that conscious supervision is important in how athletes manage and interact with their own mental processes, both prior to and during performance. This idea on metacognition also connects interestingly to a recent trend in philosophy. Several philosophers have argued that metacognition plays an important role in perception-motor expertise (Pacherie & Mylopoulos, 2021; Shepherd & Mylopoulos, 2021). The link between these philosophical positions and the psychological literature on metacognition is worth future discussion.

References


Concluding remarks and further research

Articles 1–4 have offered a line of philosophical analyses of the role of metacognition in attention control and skilled action. These articles have aimed to apply theoretical resources from philosophical psychology to illuminate the personal-level metacognitive processes of mindfulness and flow. Let me show how the analyses of Articles 1–4 prompts further philosophical discussion and research.

Article 1 argues that decentering is a puzzling mental operation, since it involves increasing the degree to which a mental state is conscious while decreasing its influence on further processing. This is not only a puzzle to the selection-for-action theory but also to theories of attention more generally (for example, the bottleneck theory, as Article 1 also proposes). I also think this puzzle applies to another prominent philosophical theory of attention, namely the structure theory of attention (STR). Article 1 goes into detail about this matter, but let me sketch the general idea. Recall that the STR claims that attention is a priority structure of an agent’s mind and that attention control is the active organizing of this structure into a fore- and background. Recall also that the STR claims that moving a mental state to the foreground means that one increases both its conscious representation and its influence on further processing (Watzl, 2017, p. 4, 147, 107, 151). Seemingly, the STR assumes that conscious representation and determination of processing are associated in this way. Given this understanding of the relationship between conscious representation and determination of processing, the phenomenon of decentering enables the following absurdum argument against the STR:

(i) When an agent, A, decenter from a mental state, M, A both increases the degree to which M is conscious (suboperation 1 of decentering) and decreases M’s determination of further processing (suboperation 2 of decentering).
(ii) Following the STR, if an agent increases the degree to which a mental state is conscious, then it means that she is moving the state more into the foreground of her own priority structure.
(iii) Following the STR, if an agent decreases a mental state’s determination of processing, then it means that she is moving the state more into the background of her own priority structure.
(iv) Following the STR, if an agent, A, decenters from a mental state, M, A both moves M more into the foreground and more into the background of her own priority structure.

There are various ways in which the STR supporter might respond to the argument. The most promising response might be to deny the account of decentering in premise (i) above. The STR might argue that decentering does not target mental states as a whole but only aspects of these states. This reply would assume a general ontological idea, namely that mental states are complexes of various aspects (sensory, affective, cognitive content, etc.), and that you can vary these independently. I am actually sympathetic to this general ontological idea, although it obviously needs to be fleshed out in greater detail if it is to offer a response to the above argument. It would also demand reformulation of the STR. In its current formulation, the STR does not think of priority structures in terms of aspects of mental states, but as in mental states as a whole (Watzl, 2017). Yet, even if we provide a satisfying elaboration of this general ontological idea, I do not think that the idea would help the STR supporter. That is, even if we assume that decentering only targets aspects of mental states, we can still present an absurdum argument against the STR. We just need to reformulate the above argument so the puzzle is that decentering moves an aspect of a mental state more into both the foreground and background. This reformulation is possible. Decentering seems to make some aspects of a mental state more conscious while also decreasing such aspects’ influence on further processing. Consider, for example, decentering from a mental state like the depressive thought “I am a failure.” This involves introspection of the cognitive content of the thought (i.e., the propositional content is more conscious through introspection) while also detaching from and decreasing the content’s influence on further processing (i.e., this content is less influential on further thinking, planning, perception, movement, etc.). This constitutes an absurdum argument as well. Decentering moves an aspect of a mental state more into both the fore- and the background of an agent’s priority structure (e.g., the cognitive content of a depressive thought). 38

Then, recall that Article 2 stressed the relevance of experimental philosophy in determining whether mindfulness meditators really endorse the mindfulness opacity hypothesis

38 The STR supporter’s most promising reply might be that priority structures come in two forms, namely structures of consciousness and structures of determination of processing. One can move mental states in these structures independently of each other. This reply solves the basic logical problem in the above absurdum argument, yet I think this is a radical revision of the STR. It is perhaps not attractive. It is beyond the present scope to discuss this matter in detail, but the revision seems to disrupt basic ideas on the relationship between phenomenal consciousness and agency in the STR (Watzl, 2017, ch. 8, 9).
Other authors have already done survey studies on expert meditators, concerning their introspective experiences relating to philosophical controversies on the nature of consciousness (e.g., Gamma & Metzinger, 2021). As Article 2 stresses, further philosophical discussion of the MOH should tackle the following question: If mindfulness meditators really accurately report that they have states of introspective awareness that match the idea of the MOH, are these states of introspective awareness driven by agents holding certain beliefs about their own experiences in advance? That is, are agents capable of opaque introspection because they believe that phenomenological properties are simply features of their experience and not external objects or properties of such objects? This is the issue of whether the opaque introspective awareness of mindfulness somehow necessitates that we have certain beliefs about the nature of our own mental states and consciousness.

I think there is a relationship between this question and the well-known philosophical discussion of cognitive penetration. Very roughly, cognitive penetration occurs when an agent’s non-perceptual states (beliefs, hopes, desires, etc.) about an object directly influence how the agent perceptually experiences the object (Siegel, 2012). This traditional form of cognitive penetration concerns an agent’s non-perceptual states about external objects and her perceptual awareness of these external objects. The present issue on mindfulness and opaque introspective awareness is different. This issue concerns whether an agent’s opaque introspective awareness of her mental state is driven by her belief about the metaphysical nature of this state. This is an issue of whether non-perceptual states about one’s own mental states (e.g., whether we believe that our mental states have nonrepresentational features) directly influence our immediate introspective awareness of these states (i.e., our direct qualitative awareness of the states). Hence, this issue of mindfulness differs from traditional cases of cognitive penetration with regard to three core aspects, namely the objects of awareness that are involved (external objects versus own mental states), the kind of awareness involved (perceptual awareness of external objects versus introspective awareness of one’s own mental states), and relevant to the kinds of non-perceptual state directly influencing awareness (non-perceptual states about external objects versus non-perceptual states about the nature of one’s own mental states). These differences are important. Hence, I propose that we label the issue relating to Article 2 and the mindfulness opacity hypothesis as metacognitive penetration. This labeling stresses the introspective nature of this issue, yet it also acknowledges the important link with cognitive penetration. Further work in philosophy might investigate the coherence, assumptions, empirical grounding, and philosophical implications of the idea of metacognitive penetration, both in relation to mindfulness and perhaps more broadly.
Article 3 examined the idea that the essential and demarcating psychological capacity of mindfulness is a form of metacognitive control. The article argued that the best way to make sense of this idea is to understand such metacognitive control as the activity of implementing certain first-order attentional or second-order attitudinal metacognitive goals. Further, the article argued that such metacognitive control is only a necessary and not a sufficient condition for exercising mindfulness. I think this account of the metacognitive control of mindfulness might inspire further discussion. Let me stress one point. I think Article 3 shows that when an agent becomes skilled in the metacognitive control of mindfulness, then she opens up new avenues of her agency, i.e., she becomes skilled in the mental action of metacognitive control. This mental action enables agents to support their own well-being, or other endeavors, in ways that do not demand direct manipulation or a change in external and environmental factors. For example, as a depressive patient, you might seek to change your mood through various actions in the environment, e.g., involvement in social interaction (you meet with friends) or motoric activity (you go for a walk). These are obviously important. Yet, the metacognitive control of mindfulness offers different routes. This control involves agents undertaking certain mental actions without overt behavior. It involves implementation of metacognitive goals. These metacognitive goals function at a proximal level (i.e., first-order attentional or second-order attitudinal metacognitive goals). Moreover, such proximal goals can be coupled with various more ultimate goals (e.g., mental health, performance, or epistemic ultimate goals).

These proposals suggest a general idea of mental action, namely that the architecture of mental action resembles the control hierarchies of bodily action. Recall, from the Introduction and Article 4, that many authors propose that bodily actions, like driving your bicycle home, involves interaction between goals at multiple levels, ranging from ultimate goals about destination, proximal goals about acute traffic, and motor goals about fine-grained motor activity such as changing gears. The explicit philosophical discussion of mental action is not as mature as that of bodily action. Yet, it seems to be a central question whether it would be correct to think that mental actions are like bodily actions in terms of control hierarchies. If this is the case, we might easily transfer insights concerning bodily actions to the domain of mental action (on topics such as skill, expertise, automaticity, etc.). If this is not the case, it is clearly interesting to grasp the differences between these two domains of human action. These questions seem worth future discussion.

Lastly, Article 4 argued that the flow argument (FA) of the automatic view (AV) was implausible in multiple respects. Hence, the article essentially makes a negative claim. Yet, I think that many of the philosophical and empirical considerations used in making this negative
claim are highly relevant in establishing more positive claims about the nature of flow. Psychologists currently disagree about how to conceptualize flow (Swann et al., 2018). Philosophers could play an important role in formulating a successful construct of flow (e.g., Shepherd, 2022). The considerations of Article 4 might be important in this endeavor. I think one core teaching of Article 4 is that rumination and worry are the phenomena that typically hinder agents from being in flow. These mental processes disrupt the deep focus needed to establish the state. I think this is why individuals with a high degree of attention control and mental toughness are more disposed to flow states. Such capacities counteract rumination and worry—they ensure that agents apply their mental resources appropriately to the relevant activity.

This not the place to present or defend my own positive account of flow. Nevertheless, I would like to mention that I support the tendency of many current psychologists in reducing the dimensions of the flow construct from nine dimensions to a lesser amount (Marty-Dugas & Smilek, 2019). Such a reduction makes the flow construct more conceptually clear and easy to operationalize in measurements. Personally, I am of the view that we should conceptualize flow as a three-dimensional construct composed of what we could call the three Fs. These are interrelated yet dissociable features of flow. They come in degrees. That is, an agent is in flow to the degree that her activity involves focus (the degree to which attention is deployed exclusively to the activity), fluency (the degree to which the activity is performed continuously without disruption or breakdown, either by the environment or by some limitation in execution), and flourishing (the degree to which the activity is meaningful to the agent). I think this conceptualization has some apparent virtues. It can make sense of why agents with identical skill levels in some activities experience different levels of flow. Importantly, it can make sense of how an agent’s psychology influences her disposition to be in flow. It depends upon psychological factors, such as the agent’s ability to focus, to remain fluent, and the degree to which she perceives and experiences the relevant activity to be meaningful. Of course, I present these claims as mere postulations here, without sufficient defense. Such defense is a topic of discussion beyond the present context.
References


