The Nature of Introspection

Adriana Renero

The Graduate Center, City University of New York

Recommended Citation
https://academicworks.cuny.edu(gc_etds/2144

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu(gc_etds

Part of the Philosophy of Mind Commons

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE NATURE OF INTROSPECTION

by

ADRIANA RENERO

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2017
The Nature of Introspection

by

Adriana Renero

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date ____________________________

Prof. Eric Mandelbaum
Chair of Examining Committee

Date ____________________________

Prof. Iakovos Vasiliou
Executive Officer

Supervisory Committee:

Prof. Jesse Prinz, Advisor
Prof. Iakovos Vasiliou
Prof. Susanna Siegel
Prof. Eric Mandelbaum
Prof. Saul Kripke

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

The Nature of Introspection

by

Adriana Renero

Advisor: Professor Jesse Prinz

My dissertation proposes a new model of introspection by examining those aspects of the nature of introspection that have been neglected in the contemporary literature, such as the ones determining variables or mental phenomena in accordance with specific cases of introspection. I assert that these neglected aspects are the very ones which provide a precise account of the way we are aware of our mental life and help us arrive at self-attributions. I begin by raising issues already extant in the epistemology of introspection, and not only argue against skeptical doubts about the reliability of introspection but also provide empirical evidence for its accuracy. I then discuss leading models of introspection, such as the inner-sense view and the acquaintance view, and show that both of these views fail to provide an explanation of the exact nature of introspection and to distinguish between different modes of introspective awareness. I finally offer a new model according to which introspection operates as what I call a cognitive superposition of mental phenomena—namely, a particular function of combining and integrating variables. It is by introspection that we become aware of the character and contents of our mental states as well as the changes, transitions, and boundaries among them. Yet while my own model of introspection builds on specific pluralist approaches which contend that introspection involves several cognitive processes, my model also shows that the operation of introspection remains distinct from the operation of other cognitive processes such as perception, attention, or inference, and thus cannot be reducible to them.
Acknowledgments

Among many debts I wish to acknowledge, let me begin with those to the Mexican National Council for Science and Technology (Consejo Nacional de Ciencia y Tecnología, CONACYT) for their three-year fellowship in support of my Ph.D. studies at The Graduate Center, City University of New York, and to The Graduate Center for its one-year Dissertation Fellowship.

It is a special pleasure to thank my principal advisor, Jesse Prinz, for his support, guidance, and encouragement at all stages of my research. From his initial belief in my project to the commitment and attention he devoted to its many stages, I greatly appreciate his patience, respect, and responsiveness, unfailingly reflected in challenges and comments of the highest caliber. He has helped me to develop my own ideas and immeasurably improved my work, teaching me along the way to be a better philosopher and better person.

I am particularly grateful for the support of my professors and dissertation committee Saul Kripke, Eric Mandelbaum, Susanna Siegel, and Iakovos Vasiliiou, as well as for their insightful feedback and constructive questions. The guidance that Susanna Siegel and Iakovos Vasiliiou have provided during my Ph.D. studies has inspired me to pay attention to what is important, and they have been ideal interlocutors. As my project has drawn to a close, their support has been essential as well.

I also want to thank my professors at CUNY’s Graduate Center, notably Noël Carroll, Stephen Neale, Gary Ostertag, Nick Pappas, Rohit Parikh, and David Rosenthal, each of whom has inspired me during my work with them. I thank, as well, professors David Chalmers and Ned Block and members of their Consciousness Group at NYU, both for their welcome and for the opportunity to examine a variety of ideas in a friendly and stimulating environment.

Parts of my work were presented at the Logic, Language, Epistemology, Metaphysics and Mind Group at Massachusetts Institute of Technology, the American Association of Mexican Philosophers at Columbia University, the Workshop of Philosophy in Spanish at Harvard University, and the Perceptual Experience Group at Universidad Autónoma Metropolitana, Mexico, I thank those audiences for helpful comments.

I owe a debt of gratitude to many colleagues. Mark Alfano, Jacob Berger, Richard Brown, Jennifer Renee Corns, Laura Di Summa, Jessica Keiser, Michal Klincewicz, Jean-David Lafrance, Rachel
Fedock, Ellen Fridland, Pete Mandik, Myrto Mylopoulos, and Dan Shargel welcomed me at the GC and I enjoyed conversations with them in the philosophy of mind and aesthetics. During these initial stages of my project I benefited from the feedback and support of Grace Helton and Lisa Miracchi, and later from the fruitful discussions and critical encouragement of Claudia Passos and Laura Pérez. I am especially grateful to Kathy Ryan for the interest and attention with which she read my work along the way, as well as for her insight into how to articulate more complex thoughts. Gina Campelia and Denise Vigani provided me with support and friendship every step of the way, and I am grateful to them for their encouragement and the wonderful experiences we shared. Thanks also to the friendship of Oliver Marshall and Romina Padró and to the Saul Kripke Center for providing me with the space and tranquility to finish the last chapter of my dissertation.

Among my Mexican professors, I wish to thank, in particular, Maite Ezcurdia, Miguel Ángel Fernández, Dora Elvira García, Olbeth Hansberg, Guillermo Hurtado, Gustavo Ortiz-Millán, Ricardo Salles, and Pedro Stepanenko for their encouragement to pursue my Ph.D. in the U.S. The list of friends and colleagues in Mexico, those who always inquired about my work, is too long to compile. Thanks, in particular, to Claudia Herrán, Jaime Carlos, Vanessa Rubio, Agustín Origüela, Alejandro Sánchez-Hidalgo, and Gabriel Sod.

Finally, I am grateful to my closest New York friends Cyrus Moore, Bev Gibbons, Jim Gorney, Denise Simon, and Paulo Vieira. My deepest gratitude is due Aslan Cohen, not only for his patience, support, and exemplary life, but above all for making my dreams come true.
# Table of Contents

**Chapter I. Addressing Skeptical Challenges to Introspection**

1. Introduction and Preliminary Considerations  
   1
2. Resisting Skepticism about the Reliability of Introspection  
   6
3. Establishing Epistemic Conditions of Introspective Access  
   12
4. Skeptical Challenges to Introspection  
   15
5. Conclusion  
   19

**Chapter II. Across-the-Board Skeptical Challenge**

1. Introduction and Preliminary Considerations  
   21
2. Sources of Error: Individual, Performance, Instruction, Interaction  
   25
3. Overcoming Obstacles: Learning, Training, and Practicing  
   29
4. Examining Introspective Judgments about Conscious Experiences  
   33
       34
   4.2. Limits are not Failures plus Grains of Introspective Judgment  
       44
   4.3. Degrees of Accuracy in Making Introspective Judgments  
       54
5. Conclusion  
   57

**Chapter III. Restrictive Skeptical Challenges**

1. Introduction and Preliminary Considerations  
   60
2. The Attitudes Challenge  
   62
   2.1. The Mindreading Account of Introspection  
       70
3. The Processes Challenge  
   76
   3.1. Equivocal Targets and Equivocal Mechanisms  
       80
4. Introspective Verbal Reports about Cognitive States and Processes  
   87
5. Conclusion  
   94

**Chapter IV. Anti-Reductionism and the Nature of Introspection**

1. Introduction and Preliminary Considerations  
   96
List of diagrams, figures, and tables

Diagram: Degrees of accuracy of introspective judgments 49
Figure 1: The “boxology of introspection” by Schwitzgebel (2012) 146
Table: The single-process model, the plural-process model, and the alternative view 149
Diagram: Scope of an introspective episode 151
Table: Mental phenomena (variables) integrated in introspection 153
Figure 2: Variables superposed (an alternative to Schwitzgebel’s view) 154
Diagram: Variables of introspection 164
Figure 3: The principle of superposition (a simple illustration) 166
Diagram: Superposition of mental phenomena 172
Table: Outcomes of the superposition theory 174
I. Addressing Skeptical Challenges to Introspection

Introspection poses one of the most important problems in epistemology, philosophy of mind, and the development of a science of consciousness. Skeptical challenges about the reliability of introspection are pressing in the philosophical and scientific debates on introspection. I argue that skeptical challenges do not defeat the hypothesis that introspection is reliable, even though appealing to introspection can sometimes lead to error. I show that we can subjectively access our minds and offer warrant for our ongoing and recently past states mental states and processes within a certain time and in a given situation. We can show that our introspective judgments can be properly connected to the facts they concern and can provide empirical evidence of the way we accurately access to our minds. In order to address skeptical challenges to introspection, I distinguish between (what I call) challenges that are across-the-board and challenges that are restrictive and maintain that neither of these challenges pose any real objection, nor are they irremediable or unable to establish boundary conditions. I propose introducing some epistemic conditions of introspective access to defuse any problem about the reliability of introspection. I discuss the problems of introspection in a number of domains and offer specific responses and solutions to its alleged failures. Chapter I offers the background for addressing the challenges, Chapter II tackles the across-the-board challenge and Chapter III tackles the restrictive challenges.

1. Introduction and Preliminary Considerations

What is required for introspection to qualify as reliable? Although robust philosophical and scientific research on introspection has recently reemerged, no consensus has been reached about the nature and reliability of that very cognitive process. Several constraints and objections are based on misconceptions, and these have unfortunately become extremely influential in the contemporary literature on introspection.

I take advantage of some philosophical contributions that have shed light on introspection and the surrounding perspectives on first-person awareness. I propose introducing some epistemic conditions of introspective access which can allow us a case by case review and help us obtain a fair picture of the issues at stake. This proposal may prove useful to researchers with a range of theoretical and empirical orientations.

I begin by drawing my preliminary approach to introspection and the requirements I consider convenient to focus on examining this phenomenon. Subsequent chapters will show what theories of introspection have to offer; both particular problems and my solutions to them will be addressed in each chapter. I shall turn to four preliminary questions that may be useful to frame the philosophical analysis of introspection. Responses to these questions can refine my initial position to it.

First, in what does introspection consist? In its simplest form, introspection is a way to recognize, to learn or to be aware of our own current and recently past mental states and processes, and to self-
attribute those states and processes. Contrary to retrospection—which is the reflection of strictly past events and experiences—introspection focuses on those concurrent states and processes as they happen in our mind or occur in the stream of consciousness which enable us to describe and report our thoughts and experiences. This distinction does not imply that some dose of short-term memory is not necessary for introspecting.

However, introspection does not necessarily result in verbally reporting our mental states or processes aloud. We may generate judgments about them, and these can be verbally reported or remain in inner speech.\(^1\) Introspective judgments and reports can also be called *upshots* or *deliverances* of introspection. The distinction between introspective judgments, introspective reports, and the introspective process itself is relevant. Judgments and/or reports may count as steps, components, or parts of the introspective process, but they are not reducible to it.

The sense of awareness at stake alludes to a faculty of self-probing our mental states and processes join together with our abilities to form judgments of our states and processes, and report them accordingly—whether a state with certain qualitative or phenomenal character can be manifested thus and so from our subjective perspective, for instance.

Even though introspection can be conceived of as type consciousness in the philosophical literature, it does not bear a distinctive phenomenology. Similarly, even though introspection can be classified as a mental process, "mental" does not bear the consciousness connotation. I use "mental" interchangeably to refer to any psychological or cognitive state.

Minimally, introspection fulfills three conditions: it is directed at one’s mind (*first-person*); it is about psychological states and processes—namely, mental entities and properties, as opposed to non-mental entities and properties (*mental*); and it is about one’s current, ongoing, and recent past mental life (*occurring* states and processes). Henceforth, these should be considered *basic conditions* of introspection.\(^2\)

Second, what kind of mental states and processes can we introspect? Arguably, we can introspect: (1) our attitudes or propositional attitudes—namely, those mental states that have both *mental*

\(^1\) I am neutral on whether judgments are or remain in inner speech. I bypass discussion on this topic.

\(^2\) Although the latter condition might be controversial, I restrict my analysis to introspection of current ongoing and recently past mental states and processes.
attitude, such as thinking, believing, desiring or wondering; and intentional content—i.e., what an intentional state is about;3 (2) our perceptions or those conscious experiences that are linked to our sensory modalities—e.g., seeing, hearing, smelling, tasting, and touching; (3) our affective states or those conscious experiences such as emotions, feelings, and bodily sensations such as pains, itches, tingles, and cramps;4 (4) our concurrent actions; and (5) our mental processes such as decision-making and goal-setting.

Now, currently ongoing and recently past mental states and processes like these (1-5) are considered targets of introspection. It is worthwhile mentioning that while judgments can stand for a kind of belief, judgments also seem to involve more aspects from our epistemic agency as well—namely, they might be more evaluative than mere beliefs. For simplicity’s sake, let us distinguish here between (1) a belief that \( p \)—a target of introspection, from (2) a judgment that \( p \)—an upshot or deliverance of introspection.

Whether we can introspectively access all types of targets or only some of them requires further examination. Likewise, whether we can introspectively access the targets in the same way or whether the same sort of cognitive-processing is involved in the introspection of this variety of targets requires examination. Notice that according to some philosophical views on the mind, different mental states might be overlapping—e.g., conscious experiences might involve attitudes, or attitudes might be consciously experienced.5 I here address mental states independently.

In the light of these remarks, however, it is important to ask what features of our mental lives can be tracked by introspection. Yet while this leads us to a further question, it is nonetheless likely that the introspective access can reach all types of mentioned targets and that the same sort of cognitive-

3 Propositional attitudes are those mental states that involve a certain relation that a subject bears towards a proposition. For instance, in believing that I am shy, I adopt an attitude of belief toward the proposition I am shy. I bypass current debates on attitudes. For simplicity’s sake we may take attitudes as mental states with intentionality. The contemporary literature also calls them cognitive states or doxastic states, e.g. beliefs; they are associated with a personal level of explanation.

4 I also use “conscious mental states” and “qualitative states” interchangeably to refer to our subjective experiences, consciousness, phenomenal consciousness, or aspects of our phenomenology. The contemporary literature also calls them non-cognitive states. A remark on “consciousness” is important here. Although consciousness is related to introspection, given that introspection underlies a type of self-consciousness, I here remain neutral to theories of consciousness and self-consciousness.

5 I here bypass the debate on the overlapping of mental states, cognitive phenomenology, or whether some phenomenal properties can be considered non-sensory.
processing is involved in the introspective access to all targets—with the caveat that different variables (mental phenomena) comprise distinct introspective cases. Such, at least, will I argue (Ch. V, §4.3.).

A common sense example of introspection might be useful here. Merely having a pain—including having a distinctive feel (the qualia) without intentionality or a mental representation of damage—does not constitute a case of introspection. Rather, consciously feeling a pain, recognizing that that is a pain, and being aware that you are in a state that hurts constitutes a typical case of introspection.

Additionally, merely looking for answers when all of a sudden I ask, for example, whether you suffer from charley-horses, does not constitute a case of introspection, strictly speaking. Rather, if we are playing tennis right now, it is a sunny day, it has been a long match, and I ask accordingly how you are feeling, it is likely that you will introspectively access the contents that are currently available in your stream of consciousness to answer my question. You might judge: "I am suffering from a terrible charley-horse..." and report, accordingly. In short, the introspective process is not necessarily triggered by any question at any time and in any context; nor does it occur all the time or "regularly," as the mere thinking process does.

This example briefly illustrates that introspection can inform us of the character of a particular mental state with its corresponding content from a subjective perspective. That is, it can convey what a particular experience is like for the individual—in qualitative terms—and what a particular experience is about—in representational terms. In addition, introspection can inform us of the singularity of that very subjective perspective or something specific about ourselves while we mentally process a particular experience.

However, it is debatable whether the introspective process is silent about the precise connection between character and content, the nature of our mental states and processes, or the exact relation among introspective targets. Nonetheless, the upshots or deliverances of introspection—i.e., the judgments and reports about our current mental states, obtained via introspection—shed light on these aspects. Presumably, introspection in tandem with other cognitive processes may also contribute to figuring out these aspects.

Third, why is the study of introspection significant? Even though introspection has a puzzling nature which sets it apart from other cognitive processes, introspection poses one of the most important
challenges in epistemology, philosophy of mind, and the development of a science of consciousness. In fact, the contemporary philosophical and scientific literature on introspection aims to provide an explanation of this mental phenomenon in conjunction with its neurophysiological function.\(^6\)

Introspection is preserved as a plausible scientific method to understand aspects of the human mind. It is significant for philosophy and cognitive sciences to develop research on emotions, beliefs, and desires, where introspective reports exhibit one’s mental life and predict behaviors such as aggression, suicide, and pro-social acts. It is also important to figure out the various aspects of one’s phenomenology, such as one’s perceptual illusions, to improve one’s cognitive capacities such as problem solving, to clinically assist in the treatment of diverse diseases and disorders such as synesthesia and autism, and being able to achieve an understanding of one’s social interactions.\(^7\)

Besides the frequent use of the introspective method in empirical research and in the philosophical theorizing about the mind, the study of introspection may also prove useful in the ordinary access of our mental lives. Given that we spend most of our waking life in permanent mental activity—even though we frequently remain unaware or indifferent to our currently ongoing and recently past mental states and processes—, we may learn to become sharper critics of our mind via introspection. It may affect positively our thoughts, purposes, choices, and actions, and be useful for our examination and understanding of further purposes, choices and subsequent actions.

Fourth, what are the principal problems that introspection faces? The main problems of introspection arise from those skeptical contemporary views that use empirical evidence, in particular, as an attempt to reduce the reliability of introspection. Philosophers and cognitive psychologists disagree among themselves whether we have a privileged access to our mental lives and whether the introspective process itself is reliable. In particular, they disagree whether the upshots or deliverances of introspection concerning conscious experiences, propositional attitudes, and mental processes are reliable or not. Skepticism seems inextricable; but interestingly, our subjective conviction of introspectively accessing our

\(^6\) Incidentally, recent findings have identified the region of the brain that controls introspection (i.e., the cortical areas); see Fleming et. al., 2010. I bypass discussion on these findings.

\(^7\) For elaboration see, respectively, Frank and Schreiber, 2013; Cytowic, 2003; Zahavi and Parnas, 2003; Butler, 2013.
mental lives seems inextricable, too. To make progress on the philosophical study of introspection requires clearing up skeptical doubts about the reliability of introspection.

These clarifications will lend support to my own model of introspection as I defend my account from skeptical challenges. They may prove useful for other accounts of introspection that are currently in development as well.

My goal in this chapter is to show that those skeptical challenges do not pose any real objection, or are remediable and thus merely require establishing boundary conditions.

The rest of this chapter has three sections that serve as background for addressing skeptical challenges. In §2, I present two scenarios of introspection and a theoretical overview of my defense for introspection's reliability. In §3, I propose establishing epistemic conditions of introspective access to defuse any problem about the reliability of introspection. I discuss reliability broadly-constructed; in particular, I allude to central views that maintain that introspection is reliable or that it is reliable in some qualified ways since we possess certain privileged access to our mental states and processes. In §4, I outline central claims or skeptical challenges to the reliability of introspection and distinguish between the across-the-board challenge and the restrictive challenges.

2. Resisting Skepticism about the Reliability of Introspection

Let us suppose that you are feeling a severe pain in your toe right now. In fact, I am stepping on it while we are standing together. Besides no doubt being in a state of physical pain, you also feel upset since I am roaring with laughter as I stomp on you. The target state of introspection concerning this scenario would thus be a conscious experience: a particular pain or a couple of pains.

Now, think of a different scenario where the target state is an attitude—or more concretely, a belief. While leaving your place, your cat is watching you as if she were crying out, “Oh, don’t leave me alone again, please!” You cannot help but be totally paralyzed, and as a result you cannot get out the door. Let us suppose that you have a corresponding belief: I am being so ridiculous for having such a relationship with my cat. This is probably a true belief, since it is the case that you are being totally ridiculous by having such an attachment to your cat. But let us set aside the truth value of your belief.
Additionally, as a consequence of your remaining motionless, it is likely that you are going to be late to the concert again. But you might not be aware of this detail.

If I ask what you are feeling or what you believe at this moment, it is likely that you are introspectively accessing the contents that are currently available in your mind or in your stream of consciousness in order to answer my question.

If I ask how you know that you are feeling that pain or having that belief, it is likely that you will express that you have a subjective conviction that—depending on the mental state you are in—you are feeling a pain in your toe, or you believe that you are being ridiculous about your relationship with your cat. Notice that in asking the question "how do you know that you are feeling that pain or having that belief?" I am asking for your grounds for judging that you are in that specific pain or have that particular belief. In other words, I am asking what justifies your holding such an introspective judgment.

If I ask whether your introspective access is trustworthy—or, alternatively, whether you trust your having these mental states without being mistaken, it is likely that you will respond with something like "yes, absolutely... Well, I am not sure I can trust introspection..."

People presume introspection is untrustworthy and are given to raising questions such as: What if there is an evil genius that can manipulate our mental states? What if our mental state is a hallucination? What if we are in such a state of chronic anxiety that it interferes with our process of introspection? Consequently, depending on the second part of your response, it is very likely that you will have in mind some of these considerations. So in order to set aside possible interferences, or what I consider pseudo-objections, it is convenient to add that we should:

- Assume that you can retain information and can recognize what you have just recently experienced, and no far-fetched scenarios are at stake.
- Assume that the target mental states are not mistaken, that your conscious states are happening—i.e., that there is no evil genius that can manipulate your mental states as he pleases, that you are not a zombie, nor are you a brain in a vat.
- Assume that you are not suffering from pathologies, that you are neither lying nor self-deceiving, that you are not hallucinating, and that you are not dreaming or having an illusion or a delusion of the
states in question. Nor are you in a state of extreme neurosis, or suffering from chronic anxiety, or under psychological stress.\(^8\)

- Assume that your true belief, if it is true, is not accidental or coincidental—i.e., no epistemic luck is involved. Moreover, there are no distractors; you are confident and can correctly apply the necessary concepts for the mental states you are having and for their contents.\(^9\) Assume that you are not simply inferring your states by watching your own behavior. You want to avoid constructing any far-fetched scenarios and instead strive to obtain the most properly functioning conditions for your introspective access.

But if no far-fetched scenarios are at stake, and if both properly functioning conditions for the introspective access and the basic conditions of introspection obtain (*first-person, mental, ongoing states*), how is it that you can be sure that you are subjectively accessing the operations of your mind, and yet believe that your introspective process is untrustworthy? Moreover, what would then make introspection trustworthy?

Anti-Cartesian and behaviorist views have negatively influenced both the philosophical and scientific study of introspection for more than a century. In the simplest terms, they have radically claimed that introspection is an unreliable method of learning about our own minds because it does not produce the same results across multiple tests. Thus, these views have stressed what they consider the failures of introspection.

No skeptical view today would deny the existence of introspection or would accept an extreme model. Indeed, with the advancement of brain-imaging techniques, interest in the benefits of introspective methodology has been re-emerging in the last decades. Recent efforts have been made by philosophers and cognitive scientists to reevaluate the positive role of introspection. But some

---

\(^8\) Although consideration of the implications of introspective process for subjects with pathologies, and subjects with conceptual or linguistic skills impaired and for children, are relevant, I restrict my research only to cases of normal adult subjects. Reyes and Sackur (2014) find that when the brain is full of cortisol or we are under psychological stress, we tend to block the process of introspection. Evidently, we can distinguish the mentioned pseudo-objections as possibilities of error or as noise according to different modalities or psychological states in cognitive psychology studies.

\(^9\) I am not assuming folk psychology here or the “ordinary person’s repertoire of mental concepts, whether or not this repertoire invokes a theory” (Goldman 1993, p. 347).
contemporary scholars still maintain certain skepticism about the reliability of introspection—henceforth, they carry a skeptical view.

Among those authors who advocate for the skeptical view are ones who claim that introspection is generally unreliable. Another set of authors restrict this claim to certain kind of mental entities. Let me call all of these authors’ claims skeptical challenges to the reliability of introspection. The former group of authors insists that introspection is usually quite unreliable across all mental conditions, including our introspective process, our mental states, and our introspective judgments and reports. These authors adopt what I will call the across-the-board skeptical challenge. The latter group of authors who refer to specific mental states such as attitudes and mental processes adopt what I will call the restrictive skeptical challenges.

I claim that we need to establish epistemic conditions of introspective access in order to defuse any problem about the reliability of introspection. I argue that skeptical challenges—both the across-the-board and the restrictive challenges—do not defeat the hypothesis that introspection is reliable, even though our appealing to introspection can sometimes lead to error. I show that we can still subjectively access our minds and can offer warrant of our ongoing and recently past states mental states and processes in a certain time and a given situation. We can show that our introspective judgments can be properly connected to the facts they concern and can provide empirical evidence of the way we accurately access to our minds. I show that skeptical challenges do not pose real objections, nor they are irreremediable or require establishing boundary conditions.

Clarifications of the approach that I defend are important. First, I do not restrict the scope of my analysis to “non-factive” mental states only. For instance, the expression “S has a belief as of p” may refer to a mental state that occurs even though p does not obtain. My analysis can include “factive” mental states: the expression “S believes that p” may refer to a true belief if the state occurs and if p obtains.10 It is worthwhile emphasizing that introspection does not necessarily have to take a position regarding the truth values of the contents that accompany our beliefs or other mental states.

Next, in my view, introspection does not presuppose the possession of knowledge of our mental states and of their processes and contents. Please bear in mind that I am not using the terms knowing or

10 I use the term factive broadly construed; for discussion on factive mental states, see Williamson 2000.
knowledge at face value. It is unlikely that introspection would qualify as knowledge. Strictly speaking, we are not always in a position to know when we are in a given mental state or process. But we can be introspectively aware of being in such a state or process. For introspection entails an awareness and self-probing of what we are processing while we access those very mental states and processes.

For the most part, philosophers tend to equate—incorrectly, I believe—self-knowledge with introspection (e.g., Shoemaker, Byrne, Carruthers, Schwitzgebel, Gertler, Smithies and Stoljar, among others). In contrast to this trend, I distinguish "introspection" from terms like "knowledge" and "self-knowledge," even though some epistemic tools respective to the problem of knowledge and justification can be adapted to analyze some aspects of introspection.

For example, we might distinguish the use of "factive" mental states—or when I succeeds in introspecting mental states that can lead to knowledge of them—from the use of "non-factive" mental states—or when I experience those states that occur even though p or my belief that p is not true—that is, when I can be introspectively aware of this state without knowing it. Alternatively, introspective judgments can entail (what I will call) qualified forms of privilege access.

Introspection is also sometimes used to describe the process of assessments about one's own mental states, irrespective of whether those assessments are epistemically justified. For simplicity's sake, I roughly distinguish 'self-knowledge' which serves as the teleological function of introspection (borrowing Aristotelian terminology: a final cause) from 'introspection' which serves as a cognitive process that may eventually lead us to get a certain knowledge of our mind or self-knowledge (borrowing from Aristotelian terminology again: an efficient cause).

Now, since the core of my approach to introspection is independent of the problem of knowledge or self-knowledge as such, I remain neutral to those theories of knowledge or theories of justification which are well-known as reliabilism—namely, the epistemic position that holds the truth-conduciveness of a belief-forming process or method.\(^\text{11}\)

Additionally, I maintain the basic conditions of introspection and the properly functioning conditions for introspective access, and I endorse the qualifications of access—-that is: context, no

\(^{11}\) I am neutral regarding reliabilism or which reliabilist theory is better or more compelling; see Goldman 2008.
generalization, the authority of error, and the possession of the right cognitive capacities. I also maintain *special conditions* for introspective access—that is: justification, attention, care, and conceptualization—without being committed to perfect forms of privilege access, as we shall see (§§3–4).

In my view, being "reliable" does not suggest being immune to error or being immune to skeptical challenges. However, recognizing error and recognizing that introspection faces skeptical challenges does not entail a quasi-skeptical view, neither does it undermine introspection *qua* process at all.

Being "reliable" instead alludes to the accurate correlation between our occurring mental states or processes and our introspective judgment or report of that mental state or process. It also alludes to the way the introspective subject can offer warrant of her ongoing and recently past mental states and processes in a certain time and a given situation. Notwithstanding that, however, my view would allow a range of reliability that could be determined according to an examination of specific cases concerning time and situation. Such an examination could provide additional useful and favorable outcomes. Upshots or deliverances of introspection would be open to revision to determine the range of reliability insofar as a new piece of evidence or further relevant information is at stake.

Finally, by claiming that skeptical challenges do not *defeat* the hypothesis that introspection is reliable, I mean to say that no argument or result provided by skeptical views override the hypothesis that introspection is reliable. In other words, skeptical views do not pose irrefutable conclusions or sufficient warrant such that the reliability hypothesis can be overridden either across all mental conditions or under particular states.

We will see that the hypothesis that introspection is reliable not only overcomes each challenge, but is also persuasive in its own right. The outcome of this view will bring about the kind of clarification of the nature and scope of introspection which may prove useful to existing theories of introspection. Additionally, it will provide a starting point for examining the nature and operation of introspection from different approaches (Ch. IV). Lastly, it will generate some bases to build a new model of introspection (Ch. V).
3. Establishing Epistemic Conditions of Introspective Access

In this section, I show that the first-person forms of awareness or privileged access to our minds cannot be dismissed as merely subjective or folk psychology. Moreover, in order to make progress on the reliability of introspection, it is necessary to establish epistemic conditions of introspective access.

Since the seventeenth century, philosophers have discussed the role of introspection in different ways. Descartes (1637/1980), for example, argued that merely having a mental state entails knowing that one is in that state. Introspective access to our mind is not only trustworthy but indubitable (Descartes, 1641/1986). Locke (1689/1975) asserted that the mind is transparent; we can get access to the inner operations of our mind via an internal sense or introspection. Hume (1739/1978) claimed that we can certainly access the basic features of our ongoing stream of consciousness.

Kant (1781/1929), on the other hand, claimed that we can come to know our mental lives just as we come to know external objects through some inner sense. Brentano (1874/1973) maintained that consciousness is accessible to the first person via a sort of “inner perception” or introspection. Husserl (1913/1982) claimed that we can surely understand our own “observations” or introspections by bracketing out previous conceptions and suspending judgments. Let us call all of these views the traditional model of introspective access.

In paying attention to the leitmotiv of the philosophical discussion, we are left with a central thesis: we have privileged access to our own minds—and in particular, to our current mental states and processes or aspects of these—by our use of a first-person method such as introspection. That is, our “epistemic position vis-à-vis propositions ascribing current mental states [or processes] to [ourselves]… is favorable [or authoritative] in a way no one else’s position is” (Alston, 1971, p. 230)—thus we have resulting an asymmetry of first-person and third-person.

Privileged access has been considered to be immune from ignorance and error involving infallibility, indubitability, or incorrigibility; thus they are understood as ideal forms of privileged access. While some philosophers have advocated for just one of these ideal forms, others have endorsed several. None are exclusive vis-à-vis the other.

For simplicity’s sake, consider the previous scenarios concerning your pain, your emotion, and your belief: where “I am in pain = p,” “I feel upset… = p”, and “I believe that I am ridiculous… = p” are
propositions that self-attribute a currently ongoing or recently past mental state, such that I arrive at a judgment wherein I am in a such-and-such mental state.

In its simplest form, when I make a judgment of \( p \), my judgment of \( p \) is \textit{infallible} if “there is no possibility of it being mistaken or otherwise prone to error.” My judgment of \( p \) is \textit{indubitable} if it “could not coherently or… consistently be [more strongly] doubted.” My judgment of \( p \) is \textit{incorrigible} “when it is impossible that it might be refuted, corrected, or otherwise improved upon,” thus ruling out any symmetry between first and third-person perspectives (Blaauw and Pritchard, 2005 pp. 81, 82, 86-7).

Although it may be controversial to consider \textit{transparency} as a form of epistemic privilege, let us also include it as a view accepted by the traditional model. In its simplest form, when I make a judgment that \( p \), my judgment of \( p \) is \textit{transparent} if \( p \) is happening in my mind; and if \( p \) is happening in my mind, I must have a true belief about \( p \). Namely, “if [I have] a mental state, then [I have] knowledge of that mental state, or at least [have] a belief to the effect that… [I have] that mental state” (Mandik, 2010, p. 125). The term transparency also has a different use in consciousness studies—namely, direct access to properties in the world. Let us take here the former sense only.

It would be useful to distinguish variations of those ideal forms of privileged access maintained by the traditional model. For twentieth-century philosophers have not simply assumed ideal forms but have endorsed reasonable qualifications. Let us call them \textit{qualified} forms of privileged access. For example, concerning \textit{infallibility}: my judgment of \( p \) is infallible by virtue of the impossibility of that judgment being mistaken within certain parameters defined by a context (Lewis, 1969; 1996). We do not self-attribute all of our mental states, since only some of them are self-presenting—i.e., by having some mental property such as feeling a tickle and considering one has it, one is then epistemically justified in self-attributing the property at issue (Chisholm, 1982).

Concerning \textit{incorrigibility}: even if I am mistaken in my judgment that \( p \), it is still incorrigible because no one else but me is in a better position to find out the truth of \( p \) (Armstrong, 1968). Concerning \textit{transparency}: if I have a mental state that \( p \), and if I possess the right cognitive capacities such as intelligence, rationality, and conceptualization, then I know that I have \( p \), and I am disposed to say "I judge that \( p \)—\( p \) is “self-intimating” (Shoemaker, 2012).
Other contemporary philosophers do not necessarily commit to forms of privileged access *per se*, but claim that introspective access to our mental states is special, provided that conditions such as attention and care are exercised by cautious subjects on their introspections of current experiences. That is, when I judge from the perspective of an attentive and careful subject on my current experience that *p* is present or that I am experiencing *p* right now, my introspective judgment is epistemically justified because it corresponds to the presence of *p* as I am experiencing it.

For example, let us say that you are experiencing a stabbing sensation and can judge it as distinct from a crawling, burning or a gnawing sensation, or you are feeling anger and you can judge it from the base of feeling that it makes your blood boil. Such a self-attribution requires that you conceptualize that particular sensation according to how the sensation appears to you (Chalmers, 1999; 2002); (Gertler, 2001; 2011).\(^{12}\)

Other contemporary philosophers account for a range of introspective judgments in a way that their trustworthiness cannot be threatened by skeptical views. Reasons for such introspective disagreement may be on account of particular situations: “individual differences, terminological variation, influence of background belief and expectation, and operational constraints.” While the first two reasons explain why sincere and competent subjects may disagree in terms of their judgments, the second two may challenge the reliability of introspection (Bayne and Spener, 2010).

The contribution of the aforementioned philosophical views is significant for underpinning any theory on the nature and reliability of introspection. Drawing on these views, I maintain that there are *qualified* forms of privileged access, or what we can call *qualifications of access*, such as (a) judging on the base of/in consideration of context, (b) no generalization of introspective target states, (c) authority of error, (d) possession of right cognitive capacities. There are also *special conditions* of introspective access, such as (e) being able to offer a justification of our ongoing and recently past mental states, (f) paying attention, (g) being cautious, (h) applying corresponding conceptualization, and consideration of particular situations provide reliability to introspection.

---

\(^{12}\) Gertler (2011) also emphasizes “paying careful attention not only to the target mental state but also to one’s judgment about that state and to the relation between these” (p. 103).
According to the success of qualifications of access plus the aforementioned special conditions of introspective access, it is reasonable to hold that introspection is reliable. Although preserving the distinction between qualifications of access and special conditions of introspective access is important, for the sake of simplicity, it would be useful to call them all epistemic conditions of introspective access. The reliability of introspection is given as a function of its different epistemic conditions.

Now, even though these conditions may be accepted and may endorse epistemic security, several skeptical challenges specific of introspection as a psychological process arise—rather than challenging first-person epistemic access that is generally a challenge about how it would be possible to know that $p$. So, we need to change the focus to empirical results. After examining skeptical challenges to introspection, I will show that these challenges do not defeat the hypothesis that introspection is reliable.

4. Skeptical Challenges to Introspection

Controversy has influenced the study of introspection since the nineteenth century and up until the late twentieth century in experimental psychology which considers non-cognitive and cognitive states as the targets of introspection. Cognitive scientists have since been focusing on whether introspection can constitute a reliable scientific method to study the human mind.\(^{13}\)

Some scientists have been relying on the introspective method as a source of scientific data. E. B. Titchener, for example, provides outlines in which the individual is able to introspect her experiences accurately. He also emphasizes that for the individual to introspect without bias, preconception, or error, it is necessary to receive sufficient introspective training (Titchener, 1901; 1910). However, the anti-Cartesian or behaviorist position denies that introspection is a reliable method of learning about our mind; it stresses what it considers the failures of introspection, claiming that introspection only yields subjective accounts that result in a lack of the objective measurements typically required by science.

Some positions also observe that introspection can sometimes modify or alter its target state, and can disrupt the ongoing event or interfere with the initial or original experience—that is, the mental state

---

\(^{13}\) Psychologists use "validity" to refer to the accuracy of a process, and "reliability" to refer to the production of the same results across multiple tests (Cf. Goldman 2004). For simplicity's sake, I use "reliability" for both.
or process that is its target (Brentano, 1874/1973; James, 1890/1981). Other positions assess this issue as unproblematic, provided that the targets of introspection are currently ongoing mental states. For example, the experience of seeing red remains unperturbed whether or not one is introspectively attending to the psychological fact that one is experiencing redness. Similar observations apply to other simple cases. When stimuli such as points of light and brief sounds are repeatedly presented to experimental participants under controlled conditions, the target states can be systematically observed without disruption (Wundt 1896/1902).

Other positions reject the reliability of deliverances or upshots of introspection—i.e., the introspective reports—that are sometimes conceived as forms of behavior. When presenting stimuli to participants and asking them to introspect their experiences, their resulting introspective reports are considered unreliable because different results are being obtained either in multiple sessions over time or across participants (James, 1890/1981; Wundt, 1912; Wundt, 1896/1902).

Since introspective reports cannot be considered sources of scientific data, behaviorists conclude that introspection is not an empirically reliable method to study the mind (Watson, 1913; 1919). It is not a reliable method, they argue, because stable patterns of behavioral responses cannot be identified (Skinner, 1945; 1957). Rather than being an immediate process, introspection is said to be an observational and inferential process prone to error (Ryle, 1949; Bem, 1972). Hence, several experimental psychologists limit the development of empirical investigations and look for more accurate methods to construct information-processing accounts without relying on introspection and its upshots or deliverances.

However, some contemporary philosophers and scientists have reevaluated the positive role of introspection. They have shown that introspection can be methodologically secure or more formally controlled using phenomenological or psychological techniques (Jack and Shallice, 2001; Jack and Roepstorff, 2002; Jack and Roepstorff, 2003; Gallagher, 2002: Lutz and Thompson, 2003; Ericsson and Simon, 1980; Ericsson, 2003).

Moreover, the psychological literature has offered interesting results about the way in which introspective reports play a reliable role in treating diseases, in personality research, emotions research, and attitudes research, where reports are predictors of important behaviors that are under people’s
conscious control, or of others behaviors such as aggression, suicide, pro-social acts, social issues, and so forth (Piccinini, 2003; Ericsson, 2003; Marcel, 2003).

Additionally, contemporary cognitive psychologists have recently offered empirical results that highlight the merits and scope of introspection, hence further advocating for the reliability of introspection. Individuals access their conscious experiences through an interactively guided introspection (Jack, 2013; Froese, 2013; Petitmengin et al. 2013). Participants can access and retrieve reliable information about their cognitive processes during visual search tasks (Reyes and Sackur, 2014).

Notwithstanding all of this, however, contemporary philosophers, still unsatisfied with the traditional model and positive reevaluations of introspection, appeal to an ignorance of recent empirical results in their insistence that introspection is unreliable, and thus preserve the skeptical view (Schwitzgebel, 2012; 2011; 2008; 2004; Carruthers, 2011; Dretske, 2012; 1995; Dennett 1991).

Permit me to present the core of the skeptical view and try to offer epistemic parallels to its corresponding claims. Introspection of our mental states and processes is unreliable; we are often mistaken, are prone to error, misrepresent our mental states and processes, and fail either when judging or in the act of reporting—which thus alludes to the fallibility of the introspective method. Moreover, because this introspection of our mental states and processes is unreliable, we cannot know that we are right about which mental state we are in or which process we are undergoing since we tend to confabulate about our mental states and processes. Hence we can be refuted or corrected in our states and processes—which thus alludes to the corrigibility of the introspective method.

Likewise, if this introspection of our mental states and processes is unreliable, we cannot identify which state we are in because we tend to distort or to modify our initial experience, so we cannot determine whether or not the content or information reported was in the original experience—which thus alluding to the non-transparency of the introspective method.\(^{14}\)

At this point, some urgent questions come to the fore in the analysis of skeptical doubts to introspection. If no far-fetched scenarios are at stake, and if basic and properly-functioning conditions obtain, how is it that you can be sure that you have subjectively accessed the operations of your mind.

\(^{14}\) For responses to the distortion challenge, see Bode, 1913; Marcel and Lambie, 2002; Marcel, 2003; Lutz & Thompson, 2003; Schooler & Schreiber, 2004; Overgaard, 2006. I bypass discussion on this problem.
and yet believe that your introspective process is untrustworthy? And what indeed would make introspection trustworthy? Be that as it may, as we have seen, a broadly construed epistemological view already holds privileged access, ideal or qualified forms, plus special conditions—i.e., epistemic conditions of introspective access—which can clear up skeptical doubts. Although we accept special conditions, one still might inquire: do we really have some qualified forms of introspective access?

A simple answer to this question may be that if we do not have certain qualified forms of access, then introspection would be reduced to other mental processes (Ch. IV discusses this problem and offers a particular anti-reductionist view). If we do have such qualified forms of access, we still need to offer an explanation of what specific kind of process is at issue—namely, we need to explain how introspection exactly operates (Ch. V discusses this problem and offers a model of the nature and operation of introspection).

We have established epistemic conditions of introspective access, yet skeptical challenges specific to introspection result from scientific investigations. So it is useful to tackle these challenges by alluding to empirical results. In other words, now, it is necessary to spell them out in psychological terms rather than in epistemic terms. Similarly, I use empirical evidence as the basis to respond to those challenges.

Let me formulate the skeptical challenges to the reliability of introspection. Recall that some authors claim that introspection is generally unreliable, while others restrict this claim to certain kind of mental entities. While the former insist that introspection is usually unreliable and highly unreliable across all mental conditions—i.e., introspective process, mental states plus introspective judgments and reports (across-the-board skeptical challenge), the latter refer to mental states such as attitudes and mental processes such as decision-making (restrictive skeptical challenges).

In the next chapter, I aim to clear up the across-the-board skeptical challenge only. It is important to recall that my objective here is not to present an account in favor of the reliability of introspection—that is, a mere contraposition to the skeptical view— but to show that the arguments that the skeptical view offers do not hold. Again, I claim that the across-the-board challenge does not defeat the hypothesis that introspection is reliable, even though appealing to introspection can sometimes lead to error. Instead of
attributing failures to the cognitive processing of introspection *per se*, I will offer specific responses and motivate solutions to the alleged failures of introspection.

5. Conclusion

In this chapter, I have framed the philosophical analysis of introspection. In particular, I have advanced considerations with regards to introspection’s *nature* (viz. brief distinctions between introspection and retrospection, and between introspection and its upshots or deliverances found in judgments and reports); introspection’s *targets* (viz. mental states such as conscious experiences and propositional attitudes, plus mental processes); introspection’s *significance* (viz. as a scientific method and ordinary access to understand aspects of one’s mental life); and introspection’s *problems* (viz. skeptical views about introspection’s reliability).

I have offered two scenarios to draw the scope and extension of introspection and have presented a theoretical landscape to resist skepticism to the reliability of introspection. Moreover, I have ruled out far-fetched scenarios and have highlighted properly functioning conditions for introspective access and basic conditions of introspection (*first-person, mental, ongoing states*).

Using different philosophical accounts in tandem, and setting the basis for the problem at stake, I have distinguished *qualifications* of access (viz. judging on the base of/in consideration of context, no generalization of introspective target states, authority of error, and possession of right cognitive capacities) from *special conditions* of introspective access (viz. being able to offer a justification of our currently ongoing and recently past mental states, paying attention, being cautious, applying corresponding conceptualization, and consideration of particular situations). I have called them all *epistemic conditions* of introspective access. I have argued that the reliability of introspection can be given as a function of their different epistemic conditions.

Finally, I have presented several skeptical views specific of introspection as a psychological process and put forward the classification of skeptical contemporary views according to two different claims: that introspection is generally unreliable (viz. *across-the-board* skeptical challenge), and that introspection of attitudes and mental processes is unreliable (viz. *restrictive* skeptical challenges).
The upshot of these considerations is not only to provide the background for addressing skeptical challenges to introspection, but also to set up boundary conditions in which introspection is successfully reliable. In the next two chapters, I will show that the alleged skeptical challenges do not threaten the reliability of introspection.
II. The Across-the-Board Skeptical Challenge

The across-the-board skeptical challenge is prominent in current philosophical and scientific debates on introspection, and may seem unanswerable. I argue that the across-the-board challenge neither presents irrefutable arguments nor has sufficient justification to override the reliability hypothesis across all mental conditions. Contrary to the challenge’s claims that introspection is generally and frequently unreliable, that it yields erroneous judgments, or even that it yields no judgments at all, I will show that no argument supposedly supported by evidence from emotion, visual perception, or cognitive phenomenology decisively proves those claims. In fact, these cases do not constitute failures of introspection to any extent. I elaborate an alternative, plausible account of introspective judgments for our conscious experiences which distinguishes between limits of introspection and failures of introspection. Moreover, rather than insisting that introspective judgments are entirely either wrong or absent, I argue that we ought to consider different grains of introspective judgment—i.e., coarse-grained and fine-grained—as well as the different degrees of accuracy of introspective judgment when judging the broad outlines of our conscious experiences and their fine-grained features.

1. Introduction and Preliminary Considerations

I take Eric Schwitzgebel’s view as the paradigmatic example of contemporary skepticism about introspection. (It is at least representative of that skeptical view.) I have called this view the across-the-board challenge since it alludes to all mental conditions.\(^{15}\) In this chapter, however, I will only analyze conscious experiences and introspective judgments about them; space prevents me from discussing other targets and deliverances of introspection.

Schwitzgebel’s recent major work (2011) argues for the unreliability of introspection concerning conscious experiences such as emotions and perceptions. Schwitzgebel also tackles cognitive phenomenology and several aspects of our phenomenology—e.g., dreams, visual perspective, visual imagery, and the echoic environment. Let us consider the latter items as strange cases for now and set them aside, since it is unlikely that dreams and the listed aspects of phenomenology can be considered targets of introspection.\(^{16}\) In contrast to cognitive phenomenology, which might also be identified as a strange or controversial case, the cases of emotion and perception that he offers are borderline cases. As we shall see, Schwitzgebel’s overall strategy is particularly inadequate in that he appeals to borderline cases to support the blanket claim that “naïve introspection” is unreliable.

Before proceeding to the discussion, it is worth mentioning that while these particular cases can be considered merely local, restrictive challenges—since they tackle conscious experiences—are

---

\(^{15}\) Although it also includes cognitive states, such as propositional attitudes and mental processes, I discuss these particular challenges in Ch. III from different views—respectively, Carruthers (in §2) and Nisbett & Wilson (in §3).

\(^{16}\) For discussion on strange cases, see Schwitzgebel, 2011, Ch. 2, 3, 4, 6, and 8.
presented by Schwitzgebel in the service of the across-the-board thesis. That is, collectively, they are
designed to show that every major kind of mental state is subject to sweeping skeptical doubts.

At the beginning of his work, Schwitzgebel exhibits a moderate skepticism about introspection. However, his work soon turns into a strong open-ended skepticism about the reliability of introspection as ordinarily executed. His central argument is revealed in the chapter entitled “The Unreliability of Naïve Introspection.” After critically examining this chapter’s challenge in the light of contemporary empirical research, I argue that his view is ultimately unpersuasive. I discuss some central problems concerning the across-the-board nature of the challenge and show the significance of recent empirical evidence that offers positive results on introspection’s reliability.

Although Schwitzgebel correctly recognizes that introspection is important, necessary, and central for both one’s life and the scientific study of the mind, he observes that we are poor introspectors in general, and goes on to argue for a much more controversial thesis. He argues that introspection is unreliable or “highly untrustworthy” in two ways: “it often goes wrong or yields the wrong result” and “it fails to do anything or to yield any result” (p. 135).

Is he here simply claiming that introspection usually does deliver indefinite or indistinct results? Or does he want to go further, maintaining that introspection does not deliver judgments or reports at all—or at any rate, not accurate ones—and so introspection generally fails? Perhaps he’s claiming that introspection qua processes might work, but its results are not relevant. If so, are the alleged failures of introspection based on the cognitive processing itself, or are they based on some particular step of the cognitive processing? It is sometimes unclear what exactly is at stake regarding the alleged failures of introspection.

Schwitzgebel also claims that even under special conditions, such as when we exercise great attention and care, “we are both ignorant and prone to error.” We often get confused or fail in assessing the causes or the processes underwriting our conscious mental states, in judging what kind of mental states we are experiencing, and in determining their corresponding labels and basic features, etc. (pp. 118-19).

Are the alleged failures of introspection, then, based on the individual’s capacity to issue judgments and generate reports, rather than on their being based on the introspective process itself or its
results? In other words, do the alleged failures of introspection result from a basic human limitation or deficiency as investigators of or inquirers into our own minds?

Several inconsistencies concerning the controversial thesis arise *prima facie*. First, claiming that introspection delivers erroneous results (or no results), whether in the form of judgments or reports, does not necessarily imply that the introspective method of accessing our own mind is unreliable. In this regard, I would say it is unlikely that a particular result, even if erroneous, can affect the *whole* introspective process and so undermine the hypothesis that introspection is generally reliable. It is (i) unlikely that the introspective process could be reduced to its stages or mere parts—i.e., to its upshots or deliverances; and (ii) unlikely that the introspective process could not provide any results at all.

To support his thesis more fully, Schwitzgebel ought to offer a more thorough analysis of specific situations in which the introspective process doesn’t work efficiently to supplement his account of those situations in which introspection fails to provide the expected result. He should rely less on one-off examples and borderline cases.

Further, if introspection were in fact “highly” unreliable, as Schwitzgebel claims, it would be necessary for him to establish criteria of measurement in order to determine why introspection is highly unreliable. Alternatively, it would be necessary that he should determine different types of error.

As for his additional claims, it is unlikely that the causes or processes underwriting our mental states can be considered *targets* of introspection, strictly speaking (cf. Ch. III, §3). (I do not wish to assert that the introspective individual could never find *some* causes of her states by undertaking a further inquiry, though, and philosophical theories of mind might indeed account for the mechanisms involved in such an inquiry.)

Furthermore, there seem to be inadequate grounds for an advocate of the "challenge" to infer that the unreliability of introspection is basically a function of the individual’s ignorance. Not only can ignorance *not* directly affect the introspective process, but it is also unlikely that it can defeat the hypothesis that introspection is reliable. If the introspective individual is self-ignorant and errs, Schwitzgebel would have to both explain how the individual fails in detail and provide a typology of errors. The notion that the individual gets wrong results because she is ignorant of how introspection works is
tantamount to claiming that the individual gets a wrong result despite being "aware" of how introspection works.

Of course error may arise in particular situations. But it is the across-the-board challenge that is untenable and too extreme. In the following sections, I will show that this challenge does not in fact pose any serious objection, for it is remediable and requires establishing boundary conditions.

Let me set out the argument for the across-the-board challenge. Since we are self-ignorant of the character and contents of our conscious experiences and of what brings us those experiences (even though our mental life seems to be important for us); and since we are prone to error in our judgment about our own current conscious experiences or in our reports of them (even though we aim to be confident in our judging); therefore, introspection is usually and highly unreliable.

An immediate reaction to this argument leads us to (§1) inquire about the sources of such self-ignorance and tendency to err; (§2) question whether self-ignorance and tendency to err constitute insurmountable obstacles for the reliability of introspection; and (§3) inquire about Schwitzgebel’s reasons for positing unreliability extensive to emotional experiences, visual perception, and cognitive phenomenology.

In inquiring on the adequacy of Schwitzgebel’s arguments, I also offer responses or motivate solutions for the problems at stake. In §1, I redirect the causes of ignorance and tendency to err; I suggest different sources of error that impute the individual according to specific situations. In §2, I show that we can learn to develop introspective capacity after receiving training and, then, by practicing which eventually lead us to predict error. In §3, I suggest a distinction between grains of introspective judgment. The result of this suggestion would bring about consideration of degrees of accuracy in making introspective judgments—i.e., when we are judging the broad outlines of our experiences and their fine-grained features.

As we shall see, trying to defeat the hypothesis that introspection is reliable, as Schwitzgebel does, is a large and questionable step. My argument not only aims to clear up skeptical doubts; it also sets the basis for defending a new model of introspection against skeptical challenges in the first place (Ch. V).
2. Sources of Error: Individual, Performance, Instruction, Interaction

In asking what the causes of ignorance and tendency are to err, Schwitzgebel may respond that such causes are based on the fact that “we never see decisive evidence of error” given that we “lack corrective feedback” and “no one ever scolds us for getting it wrong about our experience” (p. 130). However, this claim sounds unconvincing; he owes us an explanation about the necessity of corrective feedback.

In disagreeing with Schwitzgebel’s claim, I argue here that plausible sources of error are at stake. Let us call these sources of error individual, performance, instruction, and interaction errors. As we shall see, recognizing the causes of ignorance and the possibility of error does not defeat the hypothesis that introspection is reliable.

Let me start by observing that even though in some particular circumstances a third-person may be capable of interpreting some of our mental states in accurate ways—e.g., by inferring you are in pain from the fact that your toe is getting blue or by a pale expression of pain in your face—it is not clear why this third-person may be authoritative enough to scold us on the introspection of our own subjective experiences and to indicate exactly when we are in error or where the error is.

Moreover, in order to provide corrective feedback, it is unclear how a third-person could know what we are really experiencing or how exactly we are consciously experiencing \( p \) in a certain time and in a given situation—for example, whether you are experiencing pain in your toe right now instead of a numb sensation, or whether it is severe instead of a mild sensation.

It would be useful to recall epistemic conditions of introspective access that may be in play here (cf. Ch. I, §3), and that the core of introspection alludes to the self-probing faculty of our mental states working together with our own abilities to provide judgments of our mental states, and to provide reports of them accordingly. It is important to know, for example, whether the pain that you are experiencing in your toe right now can be manifested in a particular way from your subjective perspective, even if you are not totally sure that it is physiologically occurring.

Schwitzgebel and other philosophers, however, might object that the first-person access is no more reliable than the third-person’s, thus alluding to the symmetry between the first and third-person’s perspectives. Namely, we can know our minds in the same way in which we know about the minds of
others: either by inference from observation of behavior (Ryle, 1949), or on the basis of mere inference (Boghossian, 1989; Carruthers, 2011; Byrne 2011).\(^\text{17}\) Nevertheless, the symmetry between first and third-person perspectives has been already refuted (Nichols and Stich 2003). Moreover, although both inference and introspection are sources of a certain sort of knowledge, they are different cognitive processes. To equate them in terms of access or operation may lead to a reductionism of introspection to other mental processes, as I discuss in Chapters IV and V.

Let me suggest that it is likely that an introspective third-person may be capable of using a similar sort of probing inquiry both when she accesses her own mind and when she tries to understand another person’s mind. Let us call these the *modes of access*. Although her own mode of access to her mind may assist her in providing an interpretation of the other person’s current conscious experience, the first-person mode and the third-person mode will both diverge with respect to what it is like for each of them to have a painful experience in this particular time and this specific situation. This is true even if the two modes are otherwise alike in their physical and behavioral responses to pain. In other words, while the cognitive processing of introspection in the two modes likely shares a similar structure, a strict asymmetry between the third-person mode and the first-person mode still holds.

The first-person mode is authoritative in terms of the properties of an experience \(p\) or in terms of what it is like to experience thus and so, and this also includes whatever constitutes the first person’s own background and accumulation of experience. In other words, the first-person’s approach to both the character and contents of her mental states is authoritative. Let us call the approach her access to *information*. In so doing this, I am maintaining that while both first-person and third-person could share a similar mode of access, they could not share their access to information.

Now, instead of attributing the causes of our ignorance and tendency to err to an external factor or to the fact that “we lack corrective feedback and no one ever scolds us for getting it wrong about our experience,” as Schwitzgebel claims, it might be reasonable to suggest that different causes or sources of error may be in play. In considering this, it would be useful to distinguish (i) the individual’s self-boredom, indifference, or apathy towards accessing and understanding her own mind, or her disinterest in finding

\(^{17}\) I bypass discussion on the *symmetry* of first and third-person’s perspectives—i.e., whether the process by which we know our own mind is the same as the process by which we know another’s mind.
out the qualitative distinctions among her mental states; (ii) the individual’s inertia or lack of incentives to access accurately her mental life and to provide judgments or correspondent reports; or, alternatively, (iii) the individual’s inexperience in undertaking introspection or in being aware of how introspection operates and what its merits and benefits are. For simplicity’s sake, let us call this entire group of errors individual error.

Although individual error is charged exclusively to the first-person, this error can be prevented or rectified. That is, our learning about the operation of introspection can prevent us from making this kind of error. For after we learn about the operation of introspection or the procedure to undertake introspection, not only is it possible that our interest in understanding our mental lives could arise, but our individual error could also be decreased. Although the first-person mode is susceptible to error in the introspection of its mental life, this mode is potentially the most qualified and authoritative to avoid error in the introspective access to the mental life.

Although Schwitzgebel alludes to individual error when he tackles what he calls “naïve introspection,” let me set aside for the moment this kind of individual error, and bring up some sources of error concerning scientific research. It might be reasonable to suggest that additional factors could count as sources of ignorance and, as a result, account for our tendency to err. This may be seen when an individual participant in an experiment misunderstands the instructions given to appropriately perform a corresponding task after she receives a stimulus directing her to introspect her ongoing experience for the purposes of scientific research. Let us call this performance error. Alternatively, the mistake made by the individual conducting the experiment may be that he had not provided complete instructions or had provided them in an unclear or imprecise way, thus affecting the course and outcome of the participant's experience without the experimenter's being aware of the error. Let us call this instruction error.

If performance and instruction errors were what Schwitzgebel was most objecting to in his cases of concern, not only would his analysis be dramatically affected by the reports coming from the performing tasks and the contents of the participant's experience, but he would also be mistakenly calling the participant’s introspective reports unreliable. Most importantly, he would be mistakenly describing the cognitive processing of introspection or introspection qua process as untrustworthy. In order to avoid this serious (likely common) mistake of attributing unreliability to introspection, I assert that it is necessary to
analyze the specific cases more closely, and to determine what the exact source of ignorance or error is in each empirical case.

In the face of previous suggestions, one might object that since the experimenter already knows the script or series of instructions regarding the stimuli at issue, the experimenter deserves to have the authority to scold us on our introspection or to give us corrective feedback. Or at least the experimenter deserves to have the benefit of the doubt. However, some cognitive scientists might respond that sources of error are more likely due to some kind of interpersonal relation—i.e., the interaction between the experimenter and the participant, or the character of the interpretation of the participant’s reports and whether the use of interpretations are the right sorts of scientific evidence (Jack and Roepstorff, 2002). Let us call these the interaction error.

Whether interaction error in specific cases of scientific research is found or not does not seem to release us from factoring in the possibility of individual error, however. It is reasonable to suggest that we can still learn to develop introspective capacity after receiving training, after which practice can eventually lead us to predict different sorts of error. In other words, after receiving training and, then, by practicing, the individual would be able to detect and diagnose when she is in error, and in what and where the error is. Therefore, she would learn to avoid error in such-and-such cases or would have the chance to correct errors in subsequent times and situations. How exactly prediction works in introspection is a matter of a different discussion that I tackle in Ch. V.

To sum up, contrary to Schwitzgebel’s claim about the causes of ignorance and tendency to err, I have identified plausible sources of error. Namely, individual error which alludes to the ordinary subject, performance error which alludes to the particular experimental participant, instruction error which alludes to the experimenter, and interaction error which alludes to the interaction between participant and experimenter. This repertoire not only provides a suitable basis for identifying specific sources of error; it also exonerates the introspective process. Thereby, we get an unbiased designation of error.

Even though appealing to introspection can sometimes lead to a certain sort of error, identifying the proper causes of ignorance and the tendency to err is the first step in clearing up skeptical doubts on the reliability of introspection, as we have seen. In order to determine the exact causes of ignorance and tendency to err, it would be necessary to analyze specific situations (§4.1.).
So far, recognizing both the causes and possibility of error does not defeat the thesis that introspection is reliable, since the misattribution of unreliability to introspection is likely to be the real issue; and since the causes and possibility of error impute directly either the individual, the experimental individual, the experimenter or the interaction between them, rather than the cognitive processing of introspection per se.

3. Overcoming Obstacles: Learning, Training, and Practicing

Do ignorance and error constitute insurmountable obstacles for the reliability of introspection? After the previous discussion, it is possible to anticipate a response to this question—i.e., to demonstrate that ignorance and one’s tendency to err do not constitute insurmountable obstacles for the reliability of introspection. This section aims to prove that. Since we can develop and improve our introspective capacity with training and by practicing, or even learn some methodological procedures that will help us increase our introspective acuity during a certain temporal frame, we can progressively acquire introspective experience and eventually be able to predict error.

Of course, Schwitzgebel might well object to this suggestion by arguing that even after this correct form of learning and practicing and even after undergoing the process of learning for a long time, these do not preserve us from ignorance and error. He maintains this is true “despite [our possibly having] a history of introspection, despite [our possibly having] maybe years of psychotherapy or meditation or self-reflection” (p. 121). So Schwitzgebel sees no possible prospects or hopes and “no robust scientific support for such hopes” coming from a correct form of learning and practicing (p. 129). Moreover, Schwitzgebel would end up claiming that even after receiving training in order to get introspective competence, the final result would exacerbate the unreliability of introspection since alteration or modification of our initial experience may arise. This claim falls along the lines of Titchener’s psychological proposal on introspective training, and while trying to consider the possibility of advancing naïve introspection to trained introspection (Schwitzgebel 2004; 2011, Ch. 5).

At stake is not only the claim that ignorance and error do not constitute insurmountable obstacles, but also the claim that even though our introspective access is qualified only in certain ways, we can still use it to acquire an understanding of our own mental lives. And yet even if Schwitzgebel rejects all or
some epistemic conditions of introspective access, the significance of recent empirical investigations that advocate positive results on the reliability of introspection nevertheless come to the fore.

Pace Schwitzgebel, recent and robust empirical evidence shows not only that introspection is reliable, but also that learning, training, and practicing can make a significant difference in introspecting our own mental states and processes. Current investigations being done in tandem, mainly by cognitive psychologists and neuroscientists, find that we get a reliable access to our mental states and features of them through introspection.

Empirical results show that introspection can significantly improve problem-solving performance when instructions are provided to participants in order to introspect systematically. If participants are aware that they are performing a task or an action, they also tend to be aware of having an insight or of being stuck. This awareness usually leads the individual to make a decision to change her initial strategy. Indeed, new methodologies and systematic procedures have been developed for both instructing the participant to introspect and for reliably collecting introspective data (Jäkel and Schreiber, 2013).

Other empirical results show that introspection may become the most reliable measure of conscious experiences when one has been trained in it—in particular, for circumstances in which the introspection is to be “interactively guided” via an “elicitation interview.” This might occur when a method or technique of data collection is designed to gather precise descriptions of the individuals’ experiences—the kind in which the participant learns to become more introspectively aware of her conscious experiences. In these situations, when the participant is able to receive instruction and training from a suitable experimenter, the participant will more easily come into contact with her experiences and hence be better able to offer accurate descriptions of her introspection of those experiences (Petitmengin, 2006; Froese, 2013).

In fact, recent results show a significant difference between what can be concluded from the two scenarios—i.e., the one involving counting with the help of such “elicitation interview” and the one involved without it. One example can be seen in an experiment designed to reproduce the “choice blindness” paradigm (Johansson et. al., 2006; based on the research of Nisbett and Wilson, 1977) wherein the experimental participant is presented with a pair of photos of women and is asked to choose her preference. After six trials, the experimental participant is handed her chosen photo and asked to
explain her choice. In three of those trials, she is “secretly handed” the non-chosen photo; in the other three trials she is not. Interestingly, participants detected the manipulation of their choice in 80% of the cases with the help of the elicitation interview, while they detected the manipulation in only 33% without the help of the elicitation interview (Petitmengin et. al. 2013; Froese 2013). These results contribute to the significance of how differentiating circumstances contribute to the successful measuring of the reliability of her introspections.

Additional studies offer evidence of how a higher cognitive function such as introspection is more reliable when the framework of our experiences is enhanced. Proposals concerning the provision of training and the interaction between experimenter and participant—or where the former instructs the latter about her mental life during scanning—confirms that the instructions have been clearly understood by the participant and that a successful connection between the script and the experience has been determined. Additionally, methodological triangulation—in which behavioral measurements over several trials (stimulus-response), the record of brain activity (neural activity), and introspective evidence (the script of the experience and report of the experience) are registered—generate reliable results towards the scientific study of introspection (Gallagher, 2002; 2003; Jack and Roepstorff, 2002; 2003; Schooler, 2002). Studies show introspection’s reliability when it systematically co-varies with the target state, behavior, and physiological responses (Schooler and Schreiber, 2004).

In order to obtain successful results, empirical research also reveals a number of additional factors that might require control—namely, a trial’s time, its particular situation or context in which the individual is introspectively focusing on her experience, whether there are well-defined contents or unambiguous properties of the trial-taking individual’s experiences, and what the empirical models are that experimenters are using to interpret the subject’s experience. Empirical research also shows the relevance of providing the participant with training on procedure and conceptual application in order to accurately report on her experiences (Jack and Roepstorff, 2003). Additionally, empirical research shows the relevance of using the “protocol analysis method,” a process in which the participant can report,
passing through attention, without disruption of her course of thinking (Ericsson and Simon, 1980: 1993).  

Alternative proposals, such as the methodologically guided examination of participant’s experiences; the training and practicing of the phenomenological method of reduction (epoché), and the “bracketing” or setting aside opinions, presuppositions or folk psychology about conscious experiences in order to attend to what the participant is experiencing and how she is experiencing it, can also make a difference in introspection (Varela, 1996).

Neuroscientists may offer their own targeted version of systematic phenomenology by instructing their experimental participant to focus on her ongoing first-order experiences and asking her open questions while her neural and physiological data are collected. Results show that when this procedure is done, the participant is in fact able to introspect her experiences accurately, to define categories of her subjective parameters—such as her spontaneous thought process, interferences, etc.—and to report on the presence of distractions confronting her (Lutz et. al., 2002; Lutz, 2002; Gallagher, 2003).

In short, contrary to Schwitzgebel’s claim that learning and practicing do not preserve us from ignorance and error, robust empirical evidence shows that learning, training, and practicing make a significant difference in recognizing/ and reducing such ignorance and error-making and thus improve one’s introspective capacity and the obtaining of accurate outcomes. By claiming that learning, training and practice are relevant, we can assume that corresponding methodologies would provide procedures and terms of correctness. Empirical methods serve for both advancing further scientific research on this subject matter and improving ordinary introspective acuity. Finally, we learn to develop introspective capacity after receiving training and then by practicing which eventuality would lead us to predict error. Thereby, introspection can be suitably categorized as a cumulative cognitive process (see Ch. V, §4).

Even though appealing to introspection can sometimes lead to error or suggest that ignorance may be at issue, neither outcome constitutes an insurmountable obstacle for the reliability of introspection, nor does either defeat the hypothesis that introspection is reliable. We have provided

---

18 Even though Schwitzgebel (2011) is aware of the latter, he does not consider it.
19 Ch. V provides examples of how acquiring introspective acuity may be fruitful in the improvement of emotional states and in understanding cases of desires where we tend to resort to unconscious displacement.
empirical evidence of the way we can subjectively access our own minds accurately. Recognizing empirical evidence not only responds to the initial question, but it clears up skeptical doubts about the reliability of introspection.

4. Examining Introspective Judgments about Conscious Experiences

In inquiring into his reasons of defending the unreliability of introspection's being extensive to all conscious experiences, Schwitzgebel may respond that although we sincerely attempt to describe or report the state at stake in propitious situations of introspective awareness, introspection still does not work. Not only does it not yield judgments, it yields erroneous judgments, either because of “uncertainty” or “ignorance.”

So the problem of conscious experiences hinges on the problem of introspective judgment.

Let me try to sort out and formulate Schwitzgebel’s argument. He claims that we tend to confuse or fail in judging the broad outlines of our conscious experience—e.g., in judging whether a conscious experience is present or absent, or whether the experience is an itch or a tingle, or psychosomatic pain, or whether the experience changes every time that we feel a sensation or is the same. And since we tend to confuse or fail in judging the fine-grained features of our conscious experience—e.g., whether it can be located, pass rapidly or last for a while, is gross or fine, visceral or merely cognitive, entail mental imagery or proprioception—“introspection of current conscious experience, far from being secure, nearly infallible, is faulty, untrustworthy, and misleading, not just sometimes a little mistaken, but frequently and massively mistaken, about a great variety of issues” (p. 129, Cf. pp. 120-21).

Let me reiterate: it is doubtful that consideration of an error in introspective judging would justify such a strong view. Nonetheless, it is useful to build on this approach and observe that while the broad outlines of our conscious experience would involve (what we can call) occurrence and state class of the conscious experience, the fine-grained features would involve (what we can call) character, intensity, variation, and duration. It is important to be clear that Schwitzgebel neglects to mention that scope is

---

20 As opposed to Schwitzgebel, Spener (2015) correctly claims that introspective judgments under certain ordinary conditions can provide extensive, more “informative” and “scientifically respectable data” for inquiring into our conscious experiences. For discussion on Schwitzgebel (2011) see Spener, 2013. Bayne and Spener (2010) account for a range of introspective judgments whose reliability is not challenged by skeptical views. Space prevents us from discussing these arguments.
rather a matter of retrospection, and _location_ is an issue of perception. Therefore, I rule out both aspects from further analysis.

Although appealing to introspection can sometimes lead to confusion or error, I show that the across-the-board challenge does not pose any real objection and is in fact remediable if boundary conditions are established. That is, even if we were susceptible to failure in the introspective judging of the exact occurring state that we are experiencing—in its precise qualitative character, for example, or its features—the across-the-board challenge does not defeat the hypothesis that introspection is reliable. I here offer reasons to support this claim.

In order to exhibit Schwitzgebel’s overgeneralization concerning the introspection of conscious experiences, in §4.1., I discuss three cases of his, with one on emotion, one on visual perception, and one on cognitive phenomenology. I show that no argument or result provided by the across-the-board challenge to any of them proves that introspection fails to yield judgments, or yields erroneous or inaccurate judgments across all experiences.

In §4.2., I introduce principal problems on Schwitzgebel’s overgeneralization. I advance two proposals. One concerns a distinction between _limits_ and _failures_ of introspection and the other deals with _grains_ of introspective judgment and _degrees_ of accuracy in making introspective judgments—which, if adopted, not only would show that the across-the-board challenge does not defeat the aforementioned hypothesis, but it would actually lead us to dissolve some of the putative failures of introspection.

Finally, in §4.3., in order to show the proposal on degrees of accuracy at work, I use a couple of Schwitzgebel’s own examples to show how a hypothetical individual is successfully able to generate introspective judgments about the broad outlines of her conscious experiences and their fine-grained features.

### 4.1. Emotion, Visual Perception, and Cognitive Phenomenology

Let me begin by discussing the emotion case. According to the across-the-board challenge, introspection is generally unreliable; not only does it yield erroneous judgments, it yields no judgments of any sort. That is, even though we get instruction in how to introspect, when we’re asked probing questions, we get confused in introspectively judging the broad outlines of our emotional experiences. Let
me set aside for the moment the fine-grained features of these broad outlines; I just want to point out how Schwitzgebel says the across the board challenge leads us either to a state of uncertainty or to our not knowing about our emotional experiences.

Although the claim surrounding confusion and the failure to generate judgments is not limited to emotions—i.e., it could apply to visual perception and to other mental states as well—I shall follow Schwitzgebel’s examples concerning emotions only.

As we shall see, the emotion case presents several difficulties. At the very least, we can identify them as theoretical triggering, random questioning, neglecting deliverances, and neglecting limits.

The next quotations illustrate Schwitzgebel’s claims:

You have had emotional experiences, and you have thought about how they feel... If such experiences are introspectible, and if introspection is the diamond clockwork it is often supposed to be, then... you tell me: Are joy, anger, fear, and other emotional states always felt phenomenally... or only sometimes? Is their phenomenology... always approximately the same, or does it differ widely from case to case? [i.e.]... is joy sometimes in the head, sometimes more visceral, sometimes a thrill, and sometimes expansiveness, or, instead, does joy have a... distinctive... experiential character? Is emotional consciousness simply the experience of one’s bodily arousal and other bodily states...? Or... can it include, or even be exhausted by, something less literally visceral? Is emotional experience consistently located in space—[e.g.]... in the interior of one’s head and body? Can it have color? ...do we sometimes literally “see red” as part of being angry? Does it typically come and pass in a few moments... or does it tend to endure...? (pp. 120-21)

First, one might wonder whether Schwitzgebel’s case involves a problem of inattention of emotional experiences, or of working memory or conceptualization, instead of being one of uncertainty or error of introspection. Empirical evidence shows that attending visual images or bodily sensations may alter the individual’s judgment. For instance, felt pain is reduced when individuals are distracted from attending to it (Buhle and Wager, 2010: Legrain et al., 2009), and high working memory load both reduces negative affective responses to negative images (Schmeichel et al., 2008) and reduces
attentional capture by pain (Legrain et al. 2011). Ramm (2016) discusses problems of inattention, working memory limitations, and conceptual problems that may count as errors of introspection, according to Schwitzgebel. I here focus exclusively on the alleged failures of introspection; I’ve already considered attention and conceptualization, among others, as epistemic conditions of introspective access (cf. Ch. I, §2–§3).

Now, if we follow Schwitzgebel’s quotation, an initial difficulty arises in the emotional case. Let me call this difficulty theoretical triggering. Rather than being directed to the introspective individual or to occurring emotional experiences from the first-person perspective, the emotional case triggers philosophical analysis on emotional states in general or is intended to be intelligible to the theorist of emotions who can provide an explanation of the nature of the affective apparatus. Among the factors examined in the explanation would be whether emotions are always felt phenomenally, whether their phenomenology is the same, whether they have a distinctive experiential character, etc.21 But these overall questions are still too theoretical or abstract; they appeal to the theoretical framework or debate among theories of emotion instead of to introspection’s role as a first-person method.

Even though Schwitzgebel points out that it is a fact that we cannot tell by introspection whether emotions have a bodily quality, for example, the relevant point is that introspection reveals the particular feeling: introspection can inform whether my emotion has a bodily quality in this specific time and given situation. In addition, introspection can provide a report of the specific qualitative character of that emotion.

Rather than considering that introspection provides an account of the character of emotions in general, or provides an account of that which is distinctive of emotions in terms of bodily sensations in particular, the role of introspection is to inform us how a particular ongoing emotional experience appears to the individual from the first-person perspective. It is worthwhile to recall here the relevant epistemic conditions of introspective access, such as one’s judging on the base of or in consideration of context, of one’s not generalizing introspective target states, and of one’s being able to offer justification of our ongoing and recently past mental states, of paying attention, of being cautious, of applying corresponding conceptualization, and of consideration of particular situations (cf. Ch. I, §3).

21 I bypass discussion on the nature of emotions. See Prinz 2004; 2011.
Schwitzgebel expects that the introspective act will reveal not only specific feelings, but also how emotions can be defined, how they will relate to each other in terms of higher-level properties, and how they should then be interpreted. If philosophers, for example, cannot even decide on a definition of "emotion," why should we expect introspection to do so? In order to trigger introspective judgments, Schwitzgebel's questions must allude directly to the individual's currently ongoing emotional experience, to the character of her state, and to its corresponding contents according to a specific time and a given situation.

Schwitzgebel here fails in showing that introspection of our own emotional experiences cannot be revealed through introspection. But he brings up another example to get over the initial difficulty:

[M]ost of us have a rather poor sense, I suspect, of what brings us pleasure and suffering. Do you really enjoy Christmas? Do you really feel bad while doing the dishes? Are you happier weeding, or going to a restaurant with your family? ...We are remarkably poor stewards of our emotional experience... We may say we're happy—overwhelmingly we do—but we have little idea what we're talking about. (p. 121)

Even though these probing questions allude to the introspective individual, another difficulty arises for him in this case. Let me call it random questioning. It is important to be clear that merely looking for answers when all of a sudden someone asks, for example, whether you really enjoy Christmas, etc., does not strictly speaking constitute cases of introspection. Rather, if we are celebrating Christmas together right now, but if you are feeling miserable about having dinner with a bunch of dull people whom you cannot stand, and I ask you accordingly: are you enjoying Christmas? It is likely that you are going to introspectively access the contents that are currently available in your stream of consciousness in order to generate a judgment such as: No, I don't feel any connection with these people. Nor do I take any pleasure in it. This day means nothing to me. And you would be probably quite accurate in your introspective judgment of that particular feeling, though a caveat is important here: "quite" depends upon how practiced you are in introspecting your mental states.

It is worthwhile recalling that introspection is not triggered by just any kind of question at any time and in any context; nor does it occur all the time, or so regularly, as mere thinking process does. An
additional comment: the very question “do you really enjoy Christmas?” seems to include retrospection, as Christmas is an extended period of time. In other words, while Schwitzgebel is asking for an emotional generalization, it is unlikely that generalizations would be proper targets of introspection.

Again, Schwitzgebel fails in showing that introspection of our emotional experiences does not yield any introspective judgment. However, he will bring up yet another example to get over this difficulty:

My wife mentions that I seem to be angry about being stuck with washing the dishes again… I deny it. I reflect; I sincerely attempt to discover whether I’m angry—I don’t just reflectively defend myself; I try to be the good self-psychologist my wife would like me to be—and I still don’t think I’m angry. But I’m wrong, of course, as I usually am in such situations: My wife reads my face better than I introspect. Maybe I’m not quite boiling inside, but there is plenty of angry phenomenology to be discovered if I knew better how to look. (p. 123)

But here, too, an additional difficulty arises for him. Let me call it neglecting deliverances.\(^{22}\) Despite his claim that introspection does not yield judgments because of uncertainty or ignorance, pace Schwitzgebel, an introspective judgment has already been generated here. Even if it is only a simple judgment or an unexpected judgment about his emotional experience, his introspective process nevertheless has still delivered the judgment “I (still) don’t think I’m angry.” This mere introspective judgment is not only bringing about information that is shedding light on his current state; it is also giving rise to further judgments.

Moreover, even if Schwitzgebel’s wife could infer his mental state by watching his behavior, and even if she were additionally authoritative about her husband’s current state, Schwitzgebel’s judgment would still be incorrigible in a qualified form. That is, even if his introspection could not yield a judgment on the putative anger, the mere judgment “I (still) don’t think I’m angry” is incorrigible because no one else would be in a better position than he to find out what he is really feeling or whether his anger is occurring. It is he who presumes it’s not anger, and nothing prevents him from finding it out. At least, he seems to be aware of that: he himself admits that he’d probably find out “if I knew better how to look.”

\(^{22}\) There is empirical work on this issue: arguably, the famous Dutton & Aron (1974) bridge study is designed to show that we are inaccurate at introspective emotions. I bypass discussion on this study.
A final difficulty with his emotion case arises. Let me call it *neglecting limits*. Even if Schwitzgebel insists *I still don’t think I am angry*, but he also presumes *I’m wrong*; it might be the case that unconscious displacement or defense mechanisms such as repression or denial are at stake.\(^{23}\) Several states and aspects of our phenomenology that seem to be uncertain for us or make us confused, along with others that do *not* necessarily do that, might be on account of unconscious mechanisms. While these could be identified as part of the built-in limits of introspection—i.e., information is just unavailable—these are not failures of introspection. I shall turn to this point below (§4.2.). Theretofore, no argument on emotion offered by Schwitzgebel proves that introspection does not yield judgments.

I shall continue by discussing the visual perception case. According to the across-the-board challenge, introspection yields erroneous or inaccurate judgments in visual perception because we *fail* in judging the fine-grained features of our visual experiences.

Although the claim surrounding failure and generation of erroneous judgments is not limited to vision—i.e., it could apply to emotional states and to other mental states too—I here shall follow Schwitzgebel’s examples concerning visual perception and erroneous judgments.

The visual perception case presents a couple of difficulties, among them being *negligible effects* and *losing one’s bearings*.

The next quotations are meant to illustrate the claim that introspection yields erroneous judgments:

Look around a bit. Consider your visual experience as you do so. Does it seem to have a center and a periphery, differing somehow in clarity, precision of shape and color, richness of detail? ... How *broad* is that field of clarity? Thirty degrees? More? Maybe you are looking at your desk as I am. Does it seem that a fairly wide swath of the desk (a square foot?) presents itself clearly in experience at any one moment, with the shapes, colors, textures all sharply defined? (p. 125). I doubt [it]... Here is the root of the [naïve introspection’s] mistake... When the thought occurs to

\(^{23}\) Mechanisms of defense are also identified as *ego* protection in the psychological literature, e.g., we displace the original source of a negative state like “iere” to another; we rationalize to defend ourselves from negative states like anxiety; we regress to certain past immature forms of behavior; we react contrary to how we really feel; we protect ourselves from certain states and attribute them to another person; we identify with certain groups or activities that help us to set aside certain preoccupations; we repress distressing states (all of the mentioned mechanisms) and keep them unconscious.
you to reflect on some part of your visual phenomenology, you normally move your eyes (or "foveate") in that direction. Consequently, wherever you think to attend, within a certain range of natural foveal movement, you find the clarity and precision of foveal vision. It is as though you look at your desk and ask yourself: Is the stapler clear? Yes... The artificial wood grain between them and the mouse pad? Yes—each time looking directly at the object in question—and then you conclude that they all are clear simultaneously. (p. 126)

An initial difficulty in the visual perception case arises. Let me call it negligible effects. Schwitzgebel claims that to introspectively judge that my visual perception of my desk (stapler, etcetera) is rich, sharp, precise, and/or stable is entirely erroneous, since my visual perception is rather poor, indistinct, hazy, and/or shifting. Presumably, Schwitzgebel is relying on empirical results to support his claim—e.g., that my ability to detect any precise shape or color diminishes by approximately one to two degrees of arc (i.e., "about the size of your thumbnail held at arm’s length") outside the central visual field (p. 125; cf. p. 126).

Certainly, when I fixate on an object—say my desk—it may be difficult for me to maintain the image of my desk on the visual axis since movements take place not only in my eyes, but in my head, my body, and/or movement in the environment in general as well. No one would deny that our eyes will wander over the image in question; our eyes make rapid and discontinuous movements all the time. However, the mere fact that our visual phenomenology entails eye movement also makes room for marginal error, even though the effect is usually negligible for introspection.

The point at stake has to do not with the claim that identified eye movements might contribute to yield erroneous judgments, but with the claim that even though our introspective judgments may provide inexact information about the aspects of our visual perception, they are not strictly erroneous. Even if eye movement affects perception, the introspective subject can still generate introspective judgments about her visual experience in terms of grains.

Alternatively, if Schwitzgebel contends that he only is alluding to the illusion that we find detail wherever we look, we can provide another response. It is possible to avoid such a detail-heavy illusion by learning to describe vision from a single moment of gaze, rather than reporting it based on multiple
samplings or by analyzing an experience into its various constituents. I shall turn to this point below (§ 4.2.); for now, let us move on to presenting the problems that Schwitzgebel still faces.

A second difficulty arises here. Let me call it losing one’s bearings. One might wonder whether Schwitzgebel’s claim on the acuity of vision is a matter of inattention or not—in particular, whether it’s an error in attentional orienting or just an error of introspection. Several empirical studies show that individuals fail to notice unexpected events when their attention is diverted (Simons and Chabris, 1999; Simons 2000). One might fail in noticing the degree of acuity of one’s visual field because one is attending to a precise location while keeping one's eyes fixated. Recent research suggests that even if the individual was judging her visual experience as being clear all the way to the edges, Schwitzgebel might be confusing her dynamic vision with her eyes’ fixated vision. Ramm (2016) correctly observes:

> It has also been previously argued that subject’s belief in a large window of clarity is due to them making a judgment about dynamic vision in contrast to gaze fixed vision (see, Hill, 2011, p. 27; Engleburt and Carruthers, 2010, p. 251). In normal vision, the eyes perform a saccade a few times a second (e.g., Land, 1999), which the visual system combines into a single visual scene. We would not expect subjects to be aware of information at such short durations any more than we would expect them to see the images of a film as static rather than moving, or a spinning flame on the end of a pole as a point of light rather than a circle of flame. Dennett’s and Schwitzgebel’s subjects were only mistaken then if their beliefs were about fixed gaze vision. If their belief in a large window of clarity referred to typical, dynamic vision, then they (and most of us) were correct after all.

To return to my point: it is doubtful that a consideration of generations of saccades, eyes movements or eye rotations would justify the claim that introspective judgments about our visual perception are erroneous. While the function of my visual perception is to maintain fixation on the desk, the function of introspection is—among other things—to generate judgments on how that mere experience is to me as an individual. So, instead of judging how that stimulus or object per se is, or how that image exactly is—in terms of degrees of the periphery, dimensions and angles, resolution and stability, viewable wavelengths that differentiate the colors that the individual perceives, etc. —
introspective judgments bear the content and the qualitative character of the individual’s visual perception according to the way that visual perception is merely experienced from a subjective perspective.

Introspective judgments of the individual’s experience, therefore, organize that experience in terms of clarity of color, the precision of shape and color, the richness of detail, and so on. Evidently, the individual does not need to introspectively judge the stimuli or specific characteristics that perception of her desk provides. For example, it does not matter whether is she is seeing her desk as blue because it is the shortest viewable wavelength (about 380nm) or is seeing it as red because it is the longest wavelength (about 760nm), and so forth.

Notice that introspective questions ask what the individual’s grounds are for her having judged such a particular experience—i.e., what justifies the individual’s holding the introspective judgment that my visual perception of the desk (as I am seeing it right now) is clear all the way to the edges. Schwitzgebel seems to neglect that introspection will tell us how the individual is visually experiencing her desk, instead of what exactly the stimulus is or what the image of the object in question is. So far, no argument on visual perception offered by Schwitzgebel proves that introspection necessarily yields either plain or wholly erroneous judgments.

I will end this section by briefly commenting on the cognitive phenomenology case. According to the across-the-board challenge, introspection cannot deliver accurate judgments of a distinctive phenomenology of thought. This case presents, at the very least, a regular difficulty that I have identified as theoretical triggering.

The next quotations of Schwitzgebel's are meant to illustrate the claim that introspection yields erroneous judgments:

If there is such a thing as a conscious thought, then presumably we have them all the time. How could you go looking for them and simply not find them? Conversely, if there is no distinctive phenomenology of thought, how could you introspect and come to believe that there is one…? (p. 128). Think about what you plan to do after you have finished reading. Now consider: Was there something it was like to have that thought? Set aside any visual or auditory imagery you may have had. The question is, was there something further in your experience, something besides

---

24 I bypass discussion on cognitive phenomenology; see Bayne 2010; Bayne & Montague, 2011.
the imagery, something that might qualify as a distinctive phenomenology of thinking? (p. 128) … Is the answer so obvious that you can’t imagine someone going wrong about it? Is it as obvious as that your desk has drawers [or] your shirt is yellow…? [P]eople genuinely misjudge even this basic, absolutely fundamental and pervasive aspect of their conscious experience, even after putting their best introspective resources to work… (p. 129).25

A difficulty arises here. Let us call it theoretical triggering. Rather than being directed to the introspective individual’s judgment, the cognitive phenomenology case either triggers philosophical analysis on the nature of cognitive phenomenology or is meant to be answered by the theorist of cognitive phenomenology. This theorist would be able to provide an explanation of whether there is a distinctive phenomenal character accompanied by images, sensations or an immediately-experienced felt quality to thoughts or attitudes, or whether something besides that can be qualified as a distinctive phenomenology of thought. Why do we expect that introspection can do something so complex such as hold fixed auditory and visual phenomenology, judge it on the grounds of such a perceptual experience, and then decide whether there is something extra involved?

Again, Schwitzgebel’s questions appeal to theoretical debates. The theorist of cognitive phenomenology can assess the adequacy of different philosophical accounts and reach an agreement about the possibility that cognitive phenomenology is introspectively manifest or not. The debate may persist because of theoretical differences in how to interpret introspected states rather than in the examination of whether we have reliable introspective access to those states.

The lack of a cogent defense of the view that introspection produces wrong judgments about cognitive phenomenology, or fails to produce judgments in general, is problematic for the across-the-board challenge. To claim that introspection cannot respond to the demands of this strange case does not constitute a failure of introspection. I will pave the way to show that the anti-cognitive phenomenology position can be defended introspectively (Ch. V).

25 Schwitzgebel claims that introspection is less reliable than perception; I bypass discussion on this point.
To sum up, by discussing the emotion, visual perception, and briefly commenting on cognitive phenomenology cases, I have shown that no argument or result provided by the across-the-board challenge proves that introspection does not yield any judgment or that it yields erroneous or inaccurate judgments across all conscious experiences. So far, the across-the-board challenge has not defeated the hypothesis that introspection is unreliable. In what follows, I identify two principal problems with the across-the-board challenge and advance proposals on introspective judgments of our conscious experiences.

4.2. Limits are not Failures plus Grains of Introspective Judgment

There are two principal problems with Schwitzgebel’s approach: one problem concerns his concomitant notion of error and the other concerns his claim that we tend to fail in introspectively judging the broad outlines of our conscious experiences and their fine-grained features.

Although Schwitzgebel attempts to distinguish between limitations and errors at some point, it is unclear what the exact distinction is (cf. 2011, p. 135). I suggest that

(a) his notion of error is too broad; any suitable account encompassing introspective error must incorporate at least a clear distinction between limits of introspection and failures of introspection.

(b) rather than insisting that introspective judgments are either entirely wrong or absent, it should be agreed that there are different grains of introspective judgment—e.g., coarse-grained and fine-grained—that depend on many factors such as the individual’s mental abilities after training, learning, and practicing in introspection, and in the time and constancy invested in the introspection of her own experiences. This suggestion brings about consideration of degrees of accuracy in making introspective judgments.

The upshot of these proposals would contribute to understanding the alleged failures of introspection, and would clear up skeptical doubts on the reliability of introspection. After offering these proposals and analyzing specific examples, I show that empirical evidence offers interesting results restrict the aspects of our experiences to be introspected. This will make room for my suggestions on grains of introspective judgment and degrees of accuracy in making introspective judgments.
(a) Limits are not failures

There are aspects of our mental lives that involve limitations of acuteness concerning our introspective judgment. Certainly we can face difficulties while introspectively judging the contents of our mental lives. However, these limitations do not constitute failures or errors of introspection. We have reasons for thinking that Schwitzgebel is neglecting or bypassing this nuance by claiming that introspection is “highly untrustworthy” and “frequently and massively mistaken.”

That is, it would be unsound to attempt to defeat the hypothesis that introspection is reliable because of certain limitations of acuteness in judging our conscious experiences, such as emotions and their features, psychosomatic pains, or those aspects of our phenomenology which, according to the Freudian tradition, we hide from ourselves or may self-attribute when we do not really have them. This applies to hidden attitudes such as desires and beliefs, too.

Additionally, it would be unsound to attempt to defeat the hypothesis that introspection is reliable because of restrictions on access to unconscious states, or on aspects of them, or on the unconscious process of emotion, desire, or beliefs (Moran 2001; 2012). Just consider, for example, whether an unconscious process would not be in place in the washing dishes case, or in those unconscious processes involved in early visual perception and detection of phonemes (Fodor, 1983).

Moreover, it would be unsound to attempt to defeat the hypothesis that introspection is reliable because of other restrictions, such as those which concern clinical psychologists. This may include the unconscious displacement or substitution of new states for original states that may be unacceptable for the individual, or for those sometimes evident in family surroundings or social environments. These may still remain internalized as belonging to oneself, in the form of defense mechanisms, such as repression or denial, or those which may generate puzzling introspective judgments about our mental states, and/or sophisticated processes of reasoning. It is worthwhile to recall here the case of “anger” in the washing dishes’ situation.

It would also be unsound to attempt to defeat the hypothesis that introspection is reliable because of other restrictions on access concerning the causes of our mental states, processes or underlying mechanisms, which are not prima facie available through introspection. Alternatively, because of aspects
of our visual phenomenology such as detection of location—which is a matter of perception—forms of representing our body and location are properties of perception, but not of introspection.

Incidentally, other apparent restrictions of access, such as lack of confidence, stress, mental confusion or slowness, forgetfulness, issues concerning language, difficulty in concentrating, difficulty in thinking in a logical sequence, or difficulty in finding the exact words necessary to describe or report our mental states, may interfere with the introspective process or affect the deliverances of introspection. But while they might be considered limits of the introspective individual, they are not failures of introspection.

To sum up, limitations of acuteness and restrictions of access to unconscious states and their aspects, or to features concerning other cognitive processes, such as perception, attention, or memory, do not constitute failures of introspection, and neither do they defeat the hypothesis that introspection is reliable. Even though some of them may contain built-in limits—i.e., they may be places where information is just not available—and others may contain more infrangible conditions, it is possible to go through or get over some of them after receiving introspective training and by practicing, or by undertaking further probing inquiry.26

(b) Grains of introspective judgment and degrees of accuracy in making introspective judgments

There are at least three ideas for framing this proposal. First, it is reasonable to suggest that the problem of introspective judgment lies in the richness of the qualitative character of our conscious experiences—a richness which imposes a natural difficulty in the introspectively judging of both the broad outlines of those conscious states and their fine-grained features. However, if this were the case, the across-the-board challenge would not defeat the hypothesis that introspection is reliable, either. Since introspection cannot be reduced just to certain features of its target states and deliverances, the difficulty in question can in fact be overcome. Furthermore, even if we accept that some errors may be in play, we still would be able to judge our conscious experiences in grains—either coarse-grains or fine-grains or both, introspectively.

26 Even though as a consequence of being immersed in mere achievements, such as doing things, or acting in order to get external rewards, and even though we remain distracted with the non-mental domain or leave our mental life unattended, we can uncover resistances and self-defense mechanisms via introspection. Alternatively, we may eventually get a better understanding and become experts of it by practicing introspection.
Second, recall that the broad outlines of our conscious experiences would involve occurrences and state classes, while the fine-grained features of our conscious experiences would involve distinct components such as character, intensity, variation and duration (cf. §4). It is reasonable to suppose that the generation of introspective judgments is given as a function of the individual’s conscious experiences and the differences in her experiences with respect to their qualitative distinctions. More precisely, the generation of introspective judgments is a function of how many distinctions among conscious experiences the introspective individual can attain. Conscious experiences can therefore be organized by the individual’s introspective judgments; this would allow room to distinguish degrees of accuracy in making judgments.

Third, it is reasonable to suppose that our attention control and, in particular, our acquisition of experience and conceptual resources for the qualities of our conscious experiences, may affect the way in which we generate introspective judgments and offer introspective reports. Hence, depending on the amount of practice, constancy and time spent our introspective ability can become more sophisticated. Introspective judgments of conscious experiences can get more and more refined—if epistemic conditions of introspective access hold. We could get introspective competence or expertise in the process itself or only in parts of it.

Recent discussions of the role of practice and conceptual competence support my point. Although it is not specifically about introspection, there is a debate over whether experts have better perceptual acuity as opposed to having more refined concepts. Evidence exists showing how expertise contributes in perceptual judgment such as wine tasting (Ballester, et. al., 2008; Hughson, et. al., 2002), and how training and acquiring experience influence identification of odors (Laing and Francis, 1989; Livermore and Laing, 1996; Noble, et. al., 1987). Evidence also exists showing how wine novices and experts can generate different judgments (Parr, et. al., 2002; Solomon, 1990), as well as how the role of memory influences expertise (Parr et. al., 2004; Valentin, et. al., 2007).

If training and practice can make a significant difference in perceptual judgment, it is reasonable to presume that training and practice in introspection can also have an effect on introspective judgments, or so I argue.
To be clear: instead of attributing the alleged failures to introspection itself—by using the claim that introspection of our conscious experiences is “frequently and massively” unreliable—, and instead of insisting that we fail in our introspective judgment, I suggest that there are several aspects of our conscious experiences that we can introspectively judge. Consequently, we would be able to distinguish both the coarse-grains and fine-grains of introspective judgment.\(^\text{27}\)

Let me elaborate. Judging the broad outlines of a conscious experience can accord with the individual’s mental abilities after training, learning, and practicing in introspection. For example, ceteris paribus, introspectively judging between a conscious mental state \(p\) and a conscious mental state \(q\) would not be the same with an individual “A” who is able to introspectively judge the different states with their respective qualitative character, as it would be with an individual “B” who hardly distinguishes between states.

Here we would have different grains of the introspective judgment achieved by “A,” whose broad outlines of a conscious mental state in terms of occurrence and class state is distinct from that achievable by “B” who cannot so precisely judge between them. The result is that while both individuals would achieve accuracy in making introspective judgments of the broad outlines of their ongoing states, they would do so in different degrees.

Let me illustrate occurrence and class state. For simplicity’s sake, let us consider our favorite, or less controversial, scenario: the emotional state that you are undergoing while I roar with laughter when stomping on you (= \(q\)).

Consider \(q\) as experienced by an introspective individual (you), who can introspectively judge having an occurring painful physical experience (\(q'\)), which is to be distinguish from an occurring painful emotional experience (\(q''\)) with its respective qualitative character (occurrence). You may be able to differentiate \(q\) from recently past mental states or other affective reactions \(q_1, q_2, \ldots\) that you might be feeling right now—while I continue laughing—(class state). Perhaps you are immediately able to

\(^{27}\) I am neutral to phenomenal concepts here. Chalmers (2003) considers the grasp of phenomenal features as “direct.” Direct phenomenal concepts arise when an individual attends to the quality of her experience and forms concepts wholly based on the attention to the quality, thus “taking up” the quality into the concept p. 235).
compare q with other sensations of a similar type, which may tell us more about your current phenomenology.

Let me go further. Judging the fine-grained features of a conscious experience can accord with the individual’s time and constancy spent in the introspection of her own conscious mental states. For example, ceteris paribus, introspectively judging would not be the same for an individual who introspects to her conscious mental states continuously as opposed to for one who introspects to them occasionally.

Here again we would have different grains of introspective judgment of the fine-grained features of the conscious mental state of "A" in terms of character, intensity, variation and duration from that achievable by "B," who cannot so precisely judge between them—or not between all of them. The result is that both individuals "A" and "B" would achieve introspective judgment of the fine-grained features of their ongoing states, but they would achieve different degrees of accuracy in making introspective judgments.

Consider again your emotional experience q” and its fine-grained features such as character, intensity, variation and duration.

Every time that you introspectively judge a conscious experience q’ from q’, you can more precisely find their distinct qualities and fine-grained features, and thus become more sophisticated and refined in your introspective judging of conscious experience.

For instance, you can introspectively judge your emotional reaction as triggering merely disgust versus fury and judge both as distinct from being frustrated (character). You can judge whether your disgust is moderate or extreme (intensity). Also, you can judge how that emotional disturbance or how that triggered reaction varies from t₁ to t₂. That is, you can introspectively judge an immediate disgust as it occurs in your stream of consciousness and can judge how quickly it becomes anger—since I roar with laughter when stomping on you (variation). Additionally, you can judge how rapidly it passes—and may find out how it lasts in relation to similar stimuli (duration).

To illustrate the core of my suggestion, see the following diagram:
After learning, training, and practicing, broad outlines of conscious experiences may consist of:

- Occurrences
- Class state

With time spent and constancy, fine-grained features of conscious experiences may include:

- Character
- Intensity
- Variation
- Duration

Results: Different degrees of accuracy of introspective judgment

It might be reasonable to suppose that there are also target *levels* of introspective judgments—that is, what the judgment's target level of an experience is according to lower-order properties and higher-order properties, or according to the acquisition of conceptual resources. For example, a novice introspective individual may use the term *anger* to cover "rage" and "annoyance," whereas an expert introspective individual will distinguish among them. Experts seem to have more fine-grained concepts than novices do.

In addition, it might be reasonable to presume that as an introspective judgment of an experience of an unpleasant feeling (caused by the presence of danger, for example) can move from "terror," to "fear," to "apprehension," it will get more and more fine-grained. However, there seems to be a distinction between the grain of a concept and the level it picks out. Being able to judge my experience by virtue of its features (terror, fear, and apprehension) is to make more fine-grained introspective judgments. However, these features might be at the same level, whereas being able to judge introspectively my experience and move it from *healthy* anger, to *mild* anger, to *serious* anger, and to *extreme* anger may involve levels as well as start to become more and more fine-grained. Although this distinction extends my suggestions, let us set it aside here.

The same idea works for visual experiences. Again, it is reasonable to suppose that the generation of introspective judgments is given as a function of the individual's conscious experiences and the differences in her experiences with respective qualitative distinctions. More precisely, the generation of introspective judgments is a function of how many distinctions among conscious experiences the introspective individual can attain. Thus, conscious experiences can be organized by the individual's introspective judgments, and this would allow room to distinguish degrees of accuracy in making judgments.
Let us consider your visual perception of your toe as it is getting blue right now (= b), and set aside for the moment any bodily sensation. Consider that visual perception, as experienced by an introspective individual (you), who can introspectively judge between having an occurring visual experience in terms of color (b), texture (b'), and shape (b'') with their respective qualitative characters (occurrence). You may be able to differentiate b from other visual perceptions b₁, b₂... that you might be experiencing right now—while watching my amusement, for example (class state). Perhaps you are able to compare your visual experience of your blue toe with other recently past mental states or other visual perceptions of a similar type, which may tell more about your current phenomenology.

Here we would have different grains of the introspective judgment of the broad outlines of a visual perceptual experience of “A” in terms of occurrence and class state from that achievable by “B,” someone who cannot so precisely judge between them. The result is that both individuals “A” and “B” would achieve accuracy in making introspective judgments of the broad outlines of their ongoing states but achieve it in different degrees.

Notice that this is not a case of visual accuracy, but rather a case where an individual can introspectively judge about the experience in question, not about the stimulus. Again, what is relevant here is that the individual is introspectively aware of the differences in her states with respect to their qualitative characters.

Consider again your visual experience b and its fine-grained features such as character, intensity, variation and duration. Every time you introspectively judge a conscious experience b from b', you start to find their distinct qualities and fine-grained features, and you can become more sophisticated in introspectively judging conscious experiences to more refinement. For instance, you can introspectively judge your visual perception of your toe alone versus judging a visual perception of the proximity of my feet to your toe (character). You can judge whether your visual perception is of an intense blue toe or pallid yellow toe (intensity)—i.e., a blue nail or a broken nail. Also, you can judge how your visual perception of your toe varies from swollen at t₁ to deformed at t₂. That is, you can judge an immediate visual perception of your toe as it occurs in your stream of consciousness and can judge how quickly it becomes terribly inflamed (variation). Additionally, you can relatively acutely judge how quickly the swollen effect passes—and might find out how long it lasts in relation to similar stimuli (duration).
Here we would have different grains of the introspective judging of fine-grained features of a visual perceptual experience "A" in terms of character, intensity, variation and duration from that achievable by "B" who cannot so precisely judge between them. The result is that both individuals "A" and "B" would achieve introspective judgment of the fine-grained features of their ongoing states, but they each would realize different degrees of accuracy in making introspective judgments.

It might be reasonable to explore the scope of introspective judgment briefly. An expert introspective individual can introspectively judge that her visual experience has activated additional experiences. For instance, seeing me laughing not only may trigger your fury but also your desire to take revenge. Alternatively, after seeing your swollen toe, you not only may judge your experience as a stabbing sensation as opposed to a throbbing sensation, but the experience may also trigger some mental imagery of a fracture in a phalange. Additionally, introspective judgment of visually perceiving your toe can be displaced towards a different state: either a deep concern about paying a medical consultation to get an examination of your toe, or a real desire that I (the tormentor) disappear instantly. My Ch. V., §4 tackles the scope of introspection in similar cases.

Therefore, instead of claiming that we do fail in introspectively judging our conscious experiences, I have shown that there are instead grains of introspective judgment. In order to justify these grains of introspective judgment according to the broad outlines of conscious experiences, I have argued that the introspective individual should be capable of distinguishing her conscious experiences in terms at least of occurrences and state classes. In order to justify the grains of introspective judgment according to the fine-grained features of her conscious experiences, I have argued that the introspective individual should be capable of distinguishing her conscious experiences in terms of character, intensity, variation, and duration.

Although it is not easy to reach a consensus as to when the individual achieves a specific degree of accuracy in making an introspective judgment and how to measure the degree achieved, we can still admit to degrees regarding both the mental abilities of the individual after training, learning, and practicing in introspecting her conscious experiences and the time and constancy that the individual devotes to introspecting her conscious experiences. This is true even though introspecting twice as many times certainly does not guarantee twice the result or a more acute introspection.
In short, introspective judgment of the broad outlines of our conscious experiences and their fine-grained features don’t have to be the same for everybody. There could be different grains and different degrees of accuracy in making judgments, since the introspective individual would apprehend different contents, properties, and qualities of her experiences, according to her abilities, her training and practice opportunities, and the time she has to spend on it.

At this point, one might object that the above proposals are not persuasive, since they are neither based on empirical research nor do they avoid methodological difficulties of measurement when grains and degrees are involved. Yet some empirical research might make room for my proposals. In order to avoid confusion in introspective judgment, we can restrict the aspects of the individual’s experiences.

Empirical evidence shows successful results when experimenters instruct and induce individuals to deal with only one aspect of their conscious experience at a time—e.g., the intensity of a painful experience or the level of brightness in a visual perceptual experience. Indeed, some researchers require that the experimental participant offers simple responses such as button-pressing on automatic recording devices. The button-pressing is usually equivalent to responses such as it severely hurts or “the two stimuli look equally bright now” (Lambie and Marcel, 2002; Marcel, 1993).

Whether such responses can be functionally equivalent to reported speech acts involves a different discussion. The relevant point here is that the above proposals of grains of judgments and degrees of accuracy in making introspective judgments can accord with restricting aspects of our experiences.

Numerous additional current proposals offer interesting results regarding first-person data for increasing the individual’s sensitivity to her experiences and for choosing a particular occurrence of an experience to be described. Studies show a significant difference in the outcomes of introspection when experimenters provide not only instruction, but also particular conditions such as an appropriate space, time, and opportunity in order for an individual to introspect her experiences. For example, the experimenter can guide and bring the participant back to a particular experience if she moves away from it in order to provide a justification, explanation, or rationalization of her experiences. The participant is thereby able to offer accurate descriptions of her experiences, and to describe more detailed aspects of her experience progressively (Varela and Shear 1999; Petitmengin, 2009; Petitmengin, et. al. 2013).
Pace Schwitzgebel, several aspects of our subjective access are supported by brain observations, too. This may include widespread brain activation during conscious experience, high levels of regional brain metabolism in the resting state of consciousness, inner speech, visual imagery, fringe consciousness and the tip-of-the-tongue experience, executive functions of the self, and volition (Baars, 2003).

To sum up, even though we may be susceptible to fail in judging the exact occurring state that we are experiencing or its qualitative character with certain particular characteristics, the across-the-board challenge does not defeat the hypothesis that introspection is reliable. Contrary to this claim, I have shown that there are limits of introspection that should not be confused with failures of introspection. I have also suggested that the introspection of our conscious experiences would allow room to distinguish grains of introspective judgment and degrees of accuracy in making an introspective judgment about the broad outlines of conscious experiences and their fine-grained features.

4.3. Degrees of Accuracy in Making Introspective Judgments

I close this chapter by showing how the above suggestions would work for Schwitzgebel’s emotional and visual experiences. Again, instead of attributing the alleged failures to introspection, and instead of insisting that our introspective judgments are either entirely wrong or entirely absent, I here show that the introspection of our conscious experiences allows room to distinguish grains of introspective judgment and degrees of accuracy when we are judging the broad outlines of those experiences and their fine-grained features.

Let us bring back Schwitzgebel’s example about an emotional experience he has while washing the dishes. Consider that particular emotion as being experienced by an individual who can introspectively judge between occurrences of emotions—i.e., feeling anger (a) vs. not feeling anger (-a), or even feeling an apparently undetermined emotion. Since Schwitzgebel can presumably judge introspectively that his ongoing emotion is not anger, he is able to judge his state according to classes (a), (-a), and as different from other mental states, like that of a recently past feeling of shame he might have felt while his wife was bugging him or behaving in an inquisitive way. From these experiences, he generates an introspective judgment about them such as I do not think I am angry. Schwitzgebel here
would be exhibiting a certain introspective judgment of the *broad outlines* of his conscious experiences—specifically, in terms of *occurrence*.

Now in order to judge that he is not angry, it is likely that he will immediately compare this experience (-a) with other states of a similar type such as being indifferent, calm or relaxed. This very point might tell us more about his current phenomenology. Accordingly, he might generate a further judgment about his experience such as *I am not angry, but I do not like being bugged*. Notice that this judgment would be also connected to the facts that it concerns—i.e., he washes the dishes right now, he is not feeling anger, his wife continues the interrogation, etc.

Introspectively judging a conscious mental state would not be the same coming from a practiced introspective individual who is able to judge the occurrences and different state classes with their respective qualitative character, compared with a Schwitzgebel who hardly distinguishes the occurrence of his state. Yet in virtue of the mental abilities that these individuals would have acquired after receiving training, learning, and practice in the introspection of their experiences, both of these individuals would have achieved some accuracy in their introspective judgments of the broad outlines of their ongoing experience, albeit in different *degrees*.

Let us continue with Schwitzgebel’s case. Since he introspectively judges his experience as being without anger, we presume from what he expresses that he judges some of the fine-grained features of his experience in terms of character: “I’m not quite boiling inside” (Schwitzgebel, 2011, p. 123; cf. §4.1.). He can do this by introspectively judging that he is not experiencing other states with a different qualitative character, such as that of being burned by digestive fluids or of seeing stars as when feeling dizzy. Accordingly, he is able to generate a judgment such as *I do not think I am angry since I don’t feel quite boiling inside*. Schwitzgebel’s introspective judgment would exhibit a certain *degree* of accuracy about the *fine-grained features* of his conscious experience in terms of *character*.

Let us suppose that Schwitzgebel’s introspective ability to judge becomes more sophisticated after practice, so he can judge conscious experiences in a more refined way. For instance, he judges his emotional reaction as triggering mere *annoyance* versus *stress* and judges it distinct from feeling *rage* (character). He here generates a judgment such as *I don’t think I am angry; I just don’t feel excited about spending time on this annoying chore*. He introspectively judges whether his experience of annoyance is
moderate or extreme (intensity). In addition, he introspectively judges how that emotional disturbance varies from \( t_1 \) to \( t_2 \) (variation). He judges having an immediate annoyance as it occurs in his stream of consciousness and then judges how quickly it becomes tiresome, or he judges how slowly that emotion passes in relation to other emotions with similar stimuli. From here, he can generate a judgment such as

*I am being overwhelmed by such an interrogation; this feeling is troubling my nerves.*

To recall my previous point: introspective judgment of a conscious mental state would not be the same with an individual who introspectively accesses her experiences continuously and then judges its fine-grained features in terms of character, intensity, variation and duration, as Schwitzgebel does when he introspects to them occasionally and only judges an aspect of the character of his state. Depending on the individuals’ time and constancy spent in the introspection of their ongoing state, both of these kinds of individuals would achieve an introspective judgment of fine-grained features of their ongoing state, but they would achieve different degrees of accuracy of it.

The same factors may work for Schwitzgebel’s example on visually perceiving one’s desk upon whose top sits a stapler (set aside the remaining objects). Let us now explore how a hypothetical individual is able to judge this particular visual experience introspectively. Again, consider that this individual has acquired certain mental abilities in the introspection of her conscious experiences and that she devotes time and constancy in her introspections.

Consider this a very simple case of the introspection in which the individual is able to judge the broad outlines of her visual perception and some of their fine-grained features. She introspectively judges the occurrence of a visual experience of her desk with a stapler, and distinguishes this occurrence from another state class ongoing simultaneously, one perhaps having an auditory experience of a crystal-like sound. As she introspectively judges her ongoing visual perception, she judges that experience as being distinct from that of a recently past visual perception of the background, such as when she saw a great view of the city through the window. While she introspectively accesses her ongoing visual experience of the desk, she generates an introspective judgment such as *I see my desk and the stapler very clearly.* Accordingly, though the individual has formed an introspective judgment of the broad outlines of her ongoing state, we can recognize different degrees of accuracy of it.
Now, this individual can also introspectively judge the fine-grained features of her visual perception in terms of its character, intensity, variation, and duration. For example, she introspectively judges her ongoing visual experience with particular contents such as the shape of her desk: a geometric form versus an organic form or in terms of a rough texture versus a soft texture, or in terms the intensity and shape of the color vermillion color of the desk and the carmine red of the stapler. Also, she might introspectively judge how her visual perception of her desk varies from visually appearing brighter at t₁ to less bright at t₂. That is, she might judge an immediate perceptual image of his desk as it occurs in his stream of consciousness and can judge how quickly it becomes kind of dull, if not dark.

Let us add that every time that she introspectively judges her visual states, she might learn to find more fine-grained features of it, and the individual’s introspective ability might become more sophisticated when introspectively judges subsequent visual experiences. She might introspectively judge her visual experience of the desk in terms of the corresponding qualities of her experience—in particular, according to its center and periphery, shape, texture, and color, and including its properties such as its hue, saturation, and brightness. And even if we accept that some errors may be in play, the introspective individual might be able to judge roughly the broad outlines of her conscious mental states (occurrences and state classes) and some of their fine-grained features (character, intensity, variation and duration).

Accordingly, the individual’s introspective judging is not the result of a characteristic of the experiences themselves, but rather of how the introspective individual is able to self-probe her states in qualitative terms.

5. Conclusion

I have claimed that skeptical challenges do not defeat the hypothesis that introspection is reliable. After examining the central problems of the across-the-board challenge, I have shown that no argument or result provided by skeptical views override the hypothesis that introspection is reliable. After making use of empirical evidence and offering my arguments, I have further shown that this challenge does not pose irrefutable conclusions or sufficient warrant such that the reliability hypothesis can be overridden across all mental conditions and under particular states.
In this chapter, I have discussed the *across-the-board* skeptical challenge in a number of domains and have offered specific responses to the alleged failures of introspection. First, contrary to the across-the-board claim referencing the causes of ignorance and tendency to err, I have offered an account of the unbiased designation of error. I have suggested there are different sources of error that impute to the individual according to specific situations. The sources include the individual error which alludes to the ordinary features of the *subject*, the performance error which alludes to the *experimental individual or participant*, the instruction error which alludes to the *experimenter*, and the interaction error which alludes to the *interaction* between participant and experimenter (i.e., instead of imputing the cognitive processing of introspection *per se*.

Second, contrary to the across-the-board claim that learning and practicing do not protect the individual from ignorance and error, empirical evidence has been offered to show that learning, training, and practicing make a significant difference in both improving the introspective capacity and in obtaining accurate outcomes. I have suggested that this is because a cumulative cognitive process is in play; we can learn to develop an introspective capacity after receiving training and, then, by practicing, which eventually lead us to predict error.

Third, contrary to the across-the-board claim that introspection is frequently unreliable, and that introspection yields either erroneous judgments or none at all, I have shown that no argument has been yet provided—on emotion, visual perception, or cognitive phenomenology—that proves those claims. Instead, these cases do not constitute failures of introspection to any extent.

I have suggested a distinction does lie between *limits* of introspection and *failures* of introspection. Moreover, rather than insisting that introspective judgments are either entirely wrong or entirely absent, I argue that my own proposals considering coarse-grained and fine-grained introspective judgment and *degrees* of accuracy of introspective judgment are worth looking at when judging the broad outlines of our conscious experiences and their fine-grained features.

Finally, in order to show the proposal on grains of introspective judgment and degrees of accuracy in making judgments at work, I have used a couple of Schwitzgebel’s examples to show how a hypothetical individual is able to successfully generate introspective judgments about the broad outlines of her conscious experiences and their fine-grained features.
Next, we will tackle additional skeptical challenges—namely, restrictive challenges—and will fill in the details of how these challenges also do not defeat the hypothesis that introspection is reliable.
III. Restrictive Skeptical Challenges

Restrictive skeptical challenges are pressing in the philosophical and scientific debates on introspection. I argue that restrictive challenges do not present irrefutable arguments nor do they have sufficient justification to override the reliability hypothesis of introspection about both cognitive states and processes. Contrary to the claim that introspection about cognitive states—such as propositional attitudes like beliefs as opposed to non-cognitive states—is unreliable, I argue that no argument supported by evidence from cognitive states proves those claims. I show that occurrent attitudes—rather than standing attitudes—are the targets of introspection, and that introspection and interpretation are different mechanisms, with introspection being the one which holds an asymmetry between first-person and third-person perspectives. Moreover, contrary to the claim that introspection about cognitive processes such as the decision-making process is unreliable, I argue that no argument supported by evidence from these processes proves that claim, either. I instead show that psychological causes of cognitive processes—rather than the processes themselves—are erroneously considered targets of introspection, and that introspection is erroneously equated to mechanisms such as retrospection. Therefore, skeptical views end up holding (what I call) equivocal targets and equivocal mechanisms. I close by highlighting the role of introspective verbal reports play as the primary source of evidence working in favor of introspection’s reliability of both beliefs and decision-making processes.

1. Introduction and Preliminary Considerations

Several skeptical views claim that introspection is unreliable. But while some claim it is the across-the-board challenge across all mental conditions that makes introspection generally unreliable (Schwitzgebel, 2011), other authors restrict the claim that introspection is unreliable only to cognitive states and processes—namely, to certain kind of mental entities, such as propositional attitudes (which could include beliefs; Carruthers, 2011), and cognitive processes (which could include the decision-making process; Nisbett and Wilson, 1977).28

I call these precise or strict views about the reliability of introspection restrictive skeptical challenges, and plan to distinguish between the attitudes challenge and the processes challenge. It may also be useful to classify these views as referencing skepticism about cognition—contrary to skepticism about non-cognition which has been already illustrated in the examples previously discussed which mostly involve conscious experiences such as visual perception and emotion (cf. Ch. II).

The restrictive skeptical challenges to introspection can be formulated in the following way. The attitudes challenge is the skeptical view that claims that introspection about propositional attitudes (such as beliefs), as opposed to non-cognitive states (such as conscious states), is unreliable, whereas the

28 I here use "mental" and "cognitive" interchangeably and also "propositional attitudes" or "attitudes" interchangeably as instances of mental or cognitive states.
processes challenge is the skeptical view that claims that introspection about cognitive processes (such as decision-making), is unreliable.

I claim that restrictive challenges do not defeat the hypothesis that introspection is reliable, even though appealing to introspection can sometimes lead to error. I argue that these challenges either do not pose real objections or are at least remediable once boundary conditions are established. After examining restrictive skeptical challenges from a number of corners, I show that the skeptical view about cognition is unpersuasive and offer specific responses to the alleged failures of introspection.

Some aspects of the approach that I defend need reiteration here. As I have mentioned, "reliable" alludes to the accurate correlation between our occurring mental states or processes and our introspective judgment or introspective verbal report of them. Additionally, it alludes to the way the introspective individual can offer warrant of her ongoing and recently past mental states and processes within a certain time and in a given situation. I have remarked that my view allows for a range of reliability which can be determined according to an examination of specific cases concerning time and situation. Such an examination could provide additional useful and favorable outcomes, upshots, and deliverances of introspection. Moreover, should a new piece of evidence or further relevant information be introduced, introspective judgments and introspective verbal reports would be open to revision in order to determine an updated range of reliability (cf. Ch. I).

In order to examine the restrictive skeptical challenges and to fill in the details as to how these challenges do not override the hypothesis that introspection is reliable, the rest of this chapter is divided into three main sections:

§2 discusses the claim of the attitudes challenge and offers specific responses to the alleged failures of introspection. It highlights the relevance of distinguishing both occurring attitudes from standing attitudes and of interpretation from introspection, where the latter holds an asymmetry between first-person and third-person perspectives.

§3 discusses the processes challenge alluding to the decision-making process and offers specific responses to the alleged failures of introspection, showing the importance of distinguishing both the psychological causes of cognitive processes and the cognitive processes themselves (considered targets of introspection), and also mechanisms such as retrospection and reasoning from introspection.
§4 discusses the recent empirical psychological results that advocate for the role of introspective verbal reports as the primary source of evidence in favor of introspection's reliability both of cognitive states, such as the propositional attitudes found in beliefs, and cognitive processes, such as in decision-making.

2. The Attitudes Challenge

Skeptical views about cognition claim that introspection about propositional attitudes, such as beliefs, thoughts and desires, is unreliable. I call this view the attitudes challenge and claim that it does not defeat the hypothesis that introspection is reliable.

In this section, I present various views about the introspection of attitudes, and I argue that none of the objections presented by the skeptical view is in fact problematic since the targets of introspection must be considered as occurring attitudes, rather than standing attitudes, and that consequently different responses can be given, provided that the epistemic conditions of introspective access hold—i.e., judging on the basis of or in consideration of context; no generalization of introspective target states; authority of error; possession of the right cognitive capacities; being able to offer a justification of our ongoing and recently past mental states; paying attention; being cautious; and the application of corresponding conceptualization.

Before presenting the arguments for my claim, it is important to recall that propositional attitudes are those mental states that have both mental attitude—such as thinking, believing, desiring or wondering—and intentional content—i.e., what an intentional state is about. Mental states like these involve a certain relation that an individual bears towards a proposition. For instance, in believing that I am shy, I adopt an attitude of belief toward the proposition I am shy. I can self-attribute shyness or the quality of being shy, but I can also self-attribute the property of being in some mental state: to believe that I am shy and, specifically, that I am shy at this particular time and situation. The relevant cases for this discussion are self-attributions that have to do with occurring mental states only.  

---

29 For simplicity's sake we can also take "attitudes" as mental states with intentionality. I use "mental state" as a neutral term referring to attitudes such as beliefs, desires, thoughts, judgments, and so forth. Fodor (1983) divides propositional attitudes such as believing and desiring into "the representation of attitude types and the representation of attitude contents." Believing that there will be peace and desiring
Moreover, the term belief refers to an attitude I have when I consider that something is the case or something can be regarded as true—e.g., I consider it is the case that I am shy. Notice that the content of this belief concerns a subjective mental aspect: my shyness. It is relevant to highlight that the content of my introspective mental state in question either can or cannot be true. In other words, introspective states can be factive—that is, the belief occurs, and the content of such a belief is true; or it can be non-factive—that is, the belief occurs even though the content of such a belief is not true.30

That is, I do not restrict the scope of my analysis to “non-factive” mental states only. For instance, the expression "S has a belief as of p" may refer to a mental state that occurs even though p does not obtain. My analysis, however, can also include “factive” mental states, as in the expression "S believes that p," which may refer to a true belief if the state occurs and if p does obtain. Recall that introspection does not necessarily take up a position regarding the truth values of the contents that accompany our beliefs or other mental states.

Although judgments can stand for a kind of belief, judgments seem to involve more aspects from our epistemic agency than just beliefs; namely, they might be more evaluative than mere beliefs. For simplicity’s sake, let us distinguish here between a belief that p—a target of introspection—and a judgment that p—an upshot or deliverance of introspection. Additionally, some beliefs might be more ordinary or more frequent to encounter, such as those that allude to non-mental aspects or entities said to be about the world or beliefs about one’s environment—e.g., where my belief that there is a grasshopper in my taco is true if there is, in fact, a grasshopper in my taco.

With these reminders settled, we can now proceed as to what exactly the attitudes challenge contends and show that the first skeptical claim bears problems.

Several objections come to the fore in the contemporary philosophical discussion about introspection’s reliability of attitudes. Some philosophers claim that introspection about attitudes—in particular, beliefs, as opposed to conscious states—is unreliable, and offer several reasons. Roughly,

30 I borrow the term “factive” from Williamson, 2000. I use “factive” broadly construed and bypass discussion on the application of this term. For elaboration about what counts as knowledge, see Byrne, 2005.
while we can know the contents of our beliefs, we cannot know the beliefs that we take toward those contents themselves (Dretske, 2012).31

This claim, however, would not pose a problem for the reliability of introspection. Stoljar (2012) is right in observing that while this claim could undermine the achievement of knowledge, it cannot undermine its source. He takes the cognitive processing of introspection as a source of knowledge, not knowledge itself. But recall that introspection does not presuppose knowledge of our mental states in a strict sense, and introspection is not equated with knowledge or self-knowledge anyway. I have distinguished self-knowledge as being that which is understood as the teleological function of introspection from introspection as being that which is understood as a cognitive process, and one which may eventually lead us to possess or get certain knowledge of our mental life (cf. Ch. I, §2).

Other philosophers have wondered whether we actually can call up introspection of our beliefs, given that a belief is a relation between a specific piece of content and a thinker, and we cannot even yet specify that relation (Mandelbaum, 2013).

However, it is reasonable to presume that this point would entail an analysis of the nature of attitudes themselves. For whether beliefs are given as a function of specific contents or whether the relation is what the individual bears towards the contents of her beliefs, do not pose a problem for introspection, strictly speaking. It is important to be clear that introspection does not need to provide an explanation of the nature of mental states and processes—e.g., what an attitude is or how a mental process operates—even though these states and processes are considered targets of introspection. Additionally, introspection is neutral about the nature of beliefs, the process of belief formation, and what the processes are that mediate the revision, whether or not the individual is aware of them or undertakes a process of belief-revising in response to evidence that might conflict with the initial beliefs (Mandelbaum, 2013).32

---

31 Dretske (2012) is also skeptical about the thesis that the way in which we know our mind is different from the way in which we know others’ minds.

32 I bypass discussion on whether beliefs contain concepts and non-conscious experiences as their constituents (Fodor, 1998), whether beliefs are functional or dispositional patterns of response to input (Dennet, 1987), and whether beliefs are considered transparent (Byrne, 2005). I remain neutral about the nature of attitudes and conceptual differences that concern specific sorts of propositional attitudes. For the purpose of this section, my analysis on introspective access to attitudes neither tackles the attitudes-
Moreover, clinical psychologists have posed additional difficulties. They claim that we cannot get introspective access to our attitudes since both their unconscious aspects and their mechanisms interfere with attitudes in a significant way.

Certainly, unconscious displacement or defense mechanisms usually interfere with our beliefs and desires. In particular, unconscious displacement or substitution of new states for original states that are unacceptable for the individual, either by family surroundings, social environments or other reasons, nevertheless remain internalized as belonging to oneself. Additionally, defense mechanisms such as repression or denial of certain beliefs and desires bring about puzzling introspective judgments about our mental lives. Even though these claims might be considered as limits of introspection, they do not constitute failures of introspection, and limits of introspection are not challenges to introspection’s reliability, as we have seen (cf. Ch. II, §4.2).

Relatedly, experiments in social psychology have shown participants’ proclivity to hide beliefs whose contents concern religion, racism, and sexuality that are proscribed by society as unacceptable or improper. Individuals tend not to self-attribute beliefs or desires concerning those contents. This is evident even though these proscribed contents remain present in participants’ minds, or are “automatic belief-like” or “aliefs” which, though hidden, stand in conflict with explicit attitudes (Gendler, 2008). Alternatively, sophisticated processes of reasoning may also interfere with the individual’s self- attribution of attitudes concerning those issues (Reyes and Sackur, 2014).

Again, while these claims can be considered limits of introspection, limits of introspection do not pose challenges to introspection’s reliability.

On a different front, other authors have claimed that the introspection of our attitudes contains room for error which may break the relationship between our target state and our self- attribution of the state—or the judgment that I am having a certain state of p (Horgan and Kriegel, 2007). Others have claimed that there is more room for error in introspecting beliefs rather than desires, in the introspection of my desire of q, both the experience itself—a yearning feeling of q—and the self- attribution of such a desire are generated. Whereas in the introspection of my belief of p, no phenomenal character formation process, nor does it involve how beliefs respond to evidence, since these issues do not constitute a problem for introspection itself.
constitutive of the belief or no experience of belief-formation or of the belief itself is generated (Prinz, 2004).

Clearly, not all attitudes are alike. But the role of the theorist of introspection is not to offer a taxonomy of the similarities and distinctions of attitudes. Given the repertoire of attitudes on the table, however, we would be better off trying to sort them out according to the accompaniment of variables (the mental phenomena) that occur in the different cases of introspection. I reserve discussion on this topic for Ch. V. There, I will argue that the intervention of such variables is given as a function of the particular target mental state at stake and in how the introspective individual is able to discern differences in her states and to account for their respective variables.

Others might agree that we are able to introspectively access our own conscious experiences—e.g., being in toe pain, hearing a buzzing in my left ear, or feeling my heart beating from emotion—but disagree that we are able to introspectively access our own propositional attitudes given their puzzling nature or inherent difficulties that such an access entails. Contrary to conscious experiences such as emotions or perceptions, or even to attitudes such as desires, beliefs do not seem to entail a phenomenal character or what something is like to be in them. Hence, some philosophers argue that we may end up in error while introspectively judging whether we have a certain belief. This makes it difficult for us to rely on an introspective access to our beliefs.

To deal with such a difficulty, some authors have contended that there is indeed a certain phenomenal character of beliefs. In particular, advocates of cognitive phenomenology claim that there is a sort of phenomenology of beliefs that is not solely reducible to conscious experiences (Bayne and Montague, 2011). So we could end up having introspective access to both conscious experiences and attitudes in a similar if not same way. Although arguing in favor of a cognitive phenomenology view might be useful to clear up skeptical doubts about the reliability of an introspection of beliefs, mental states are addressed independently in this project. Most importantly, consideration of a certain overlapping of mental states brings about additional implications. I do not pursue this problem here (see Ch. V.).

An alternative response to the problem above is that even if there is no phenomenal character of beliefs themselves, there is some phenomenal aspect accompanying our beliefs. Some authors argue that beliefs, in particular, entail certain “auditory images of sentences.” “We gain conscious access to
beliefs by figuring out what we are disposed to say,” even if those images remain in “subvocal speech” (Prinz, 2004, pp. 54-55). However, whether or not beliefs entail either a phenomenal character or a phenomenal aspect accompanying them in a certain way is not an issue that the theorist of introspection should solve.

One might wonder whether it is plausible that we form an imagery of our beliefs before having the beliefs themselves. Auditory imagery or inner speech can certainly arise in the belief-formation process, or it can accompany beliefs in a certain way.

It is reasonable to presume that auditory imagery contributes in such a formation process or can appear as such before we have a belief. But a belief could also follow auditory imagery or get feedback from either an individual instance or a combination of all the instances. Since the order in which these occur may be negligible, some people might presume that beliefs and imagery match or overlap or that the phenomenal aspect is necessary for beliefs, or that beliefs can be experienced. Additionally, they may presume that imagery and belief are going to match in terms of content, and that we would be unable to tell them apart subjectively. Again, whether or not beliefs entail a phenomenal character or a phenomenal aspect accompanying them in a certain way is not an issue that the theorist of introspection should solve.

However, it is necessary to set aside this consideration, since auditory imagery does not seem to need to be constitutive of beliefs in order to be introspected. Nor does introspection seem to depend on any phenomenal character or aspect constitutive of its target beliefs, or entail any phenomenology such that there is something it is like to introspect our mental states. In the next chapter, I will discuss theories of the nature of introspection.33 In the meantime, let me suggest that there is no problem in introspectively accessing our beliefs, provided that these are occurrent beliefs or beliefs that lack constancy. I maintain that we can be introspectively aware of our beliefs by being in the relevant cognitive state; that is, that while we are self-probing those states, we are having an occurrent belief to the effect that we are standing for something, are taking a position, or are sustaining a position under certain conditions.

33 I here refer to "introspective access" only, without advocating for a direct or immediate access. I do reserve specific discussion on theories of direct introspection and causal introspection in Ch. IV.
Let us bring back one of our initial scenarios (cf. Ch. I, §2). While leaving your place, your cat is watching you as if she were crying out: “Oh, don't leave me alone again, please!” You can't help it; you get totally paralyzed and can't get out the door. Let us suppose that you have a corresponding belief: I am so ridiculous for having such a relationship with my cat. Probably, it is a true belief, since it is the case that you are being totally ridiculous by having such an attachment to your cat. But we may set aside the truth-value of your belief. Also, as a consequence of your remaining motionless, it is likely that you are going to be late to the concert again. But you might not be aware of this detail.

What if I ask you what you believe at this precise moment? It is likely that you will introspectively access the contents that are currently available in your mind in order to answer my question and produce a judgment. If I ask you how you know that you have that belief, it is likely that you will say it’s your subjective conviction that you believe you are being ridiculous about your relationship with your cat. After all, it is you and no one else who stands for that belief with that particular intentional content, and forming an introspective judgment accordingly.

If I ask you whether your introspective access is trustworthy, it is likely that you will be ready to respond with something like: “Yes, I trust I am having that belief and am not mistaken about it.” After all, that belief is telling you something about you, and you are the one best able to offer warrant of your current mental state at this precise moment and in this given situation. You will thus be ready to respond based on properly-functioning conditions for the introspective access—provided, of course, that no far-fetched scenarios are at stake.

Moreover, if basic conditions of introspection obtain, you are able to access the operations of your mind from a first-person perspective. It is worthwhile to recall that introspection fulfills the following conditions. At the very least, it is directed at one’s own mind (first-person). And it is about psychological states and processes—namely, mental entities and properties, as opposed to non-mental entities and properties (mental). Finally, it is about one’s current, ongoing, and recently-past mental life (ongoing states and processes).

Further, since epistemic conditions of such an access obtain, you can trust introspection. That is, your introspective judgment about your belief of being ridiculous is not mistaken based on the context—
i.e., that you got totally paralyzed, that you were unable to get to the door, and that it is likely that you are having a co-dependent relationship with your pet.

Additionally, you are self-attributing the current belief you are having right now about your own ridiculousness. Set aside whether the content of your belief can be defined as true or not. No generalization of target mental states is occurring: you are not self-attributing to this belief all mental states, but only some of them. You have a belief, you stand for it and you consider yourself to be having this particular belief.

And even if you had been mistaken about the content of your belief, you are nevertheless epistemically justified in self-attributing the belief at play. Suppose, for example, that you are not being ridiculous, or that someone else has already judged you as not being ridiculous. Still, it is you who possess the right cognitive capacities and are in the better position to find out whether that belief may be a true belief. Like it or not, it may be the case that you are being totally ridiculous by having such an attachment to your cat.

Finally, we can say that your introspective judgment about such a belief is justified. In particular, that your judgment *I am so ridiculous for having such a relationship with my cat* corresponds to the presence of the belief in question proves that belief is occurring, and that you are able to self-attribute that specific belief with that particular content insofar as you apply the appropriate conceptualization for both the state in question and its content. Recall that introspection focuses on concurrent states and processes as they occur in our mind, thus enabling us to describe and report our thoughts and experiences. It does not imply that a dose of short-term memory is unnecessary for introspecting. The sense of awareness at stake alludes to a faculty able to self-probe our mental states and processes along with our abilities to form judgments of our own states and processes, and to report them all accordingly.

To summarize: in the light of a broadly construed epistemological view holding that qualified forms of privileged access plus special conditions—i.e., the *epistemic conditions* of introspective access—convinces you that your introspective access is reliable, being reliable still does not suggest being
immune to error. Notice that recognition of error and recognition that introspection faces skeptical challenges does not make you a skeptic or a quasi-skeptic at all.  

So far, then, even though there is room for error, the claim that introspection of beliefs, as opposed to introspection of conscious states, is unreliable would not pose a problem for the reliability of introspection, provided that the conditions above obtain. Most importantly, the key point of this discussion is to assert that the targets of introspection are *occurent* attitudes instead of *standing* attitudes.

Propositional attitudes *qua* suitable targets of introspection allude to *occurent* mental states—beliefs, thoughts, desires, and so on. More precisely, they allude to occurrent states with explicit propositional contents: I believe that *p*, I think that *p* or I desire that *q* explicitly (for example: *p* = *I am shy of telling you what I am thinking now* and *q* = *to be in another planet right now*). It is commonly agreed that by having an attitude, a representation with an explicit propositional content is actually present in my mind, pointing both to its being such-and-such the case, and to its being such-and-such the case for the very same individual having the representation. Let us here be neutral as to what exactly "representation" accounts for.

Since the targets of introspection are occurrent mental states with explicit propositional contents, it is reasonable to maintain that introspection would exclude the so-called *dispositional states*, or states that are about things that are prone to occur or what the contemporary literature on beliefs identifies as *standing states*—roughly, states that do not play any role in our mind in the moment. As a matter of fact, these kinds of mental states are frequently examined in experiments on social psychology and serve as causal explanations for other mechanisms or cognitive processes such as reasoning.

2.1. The Mindreading Account of Introspection

As we have seen, several difficulties come to the fore in the contemporary philosophical debate on introspection of attitudes, and several responses to the alleged failures of introspection have been offered. The problem defuses if we take it that the targets of introspection regarding attitudes are

---

34 See background for addressing skeptical challenges to introspection. Now you are convinced that your introspective access is reliable, as opposed to the initial scenario in which you were not sure whether introspection is trustworthy or not. Using different philosophical accounts in tandem, and setting the bases for the problem at stake, I have set *epistemic conditions* of introspective access (cf. Ch. I, §3).
occurrent attitudes. But none of these difficulties has been so overarching in attempting to undermine the reliability of introspection about these cognitive states than Carruthers’ view has been.

In this section, I argue that the attitudes challenge does not defeat the hypothesis that introspection is reliable. That is, no argument or result provided by Carruthers’ view in *The Opacity of Mind* (2011) poses irrefutable conclusions or sufficient warrant such that the reliability hypothesis can be overridden under particular cognitive states. It is unlikely that introspection is the very cognitive process at the core of Carruthers’ view. In other words, it is doubtful that Carruthers’ theory can be considered a theory of introspection in the first place. Trying to accommodate the first-person perspective within the framework of the third-person perspective not only implies neglecting basic conditions of introspection, but also shows that his view is unpersuasive.

In order to show that the attitude challenge does not defeat the hypothesis that introspection is reliable, let us begin by pointing out some of the basics of Carruthers’ theory. Let us then discuss the implications of accepting this view.

Although Carruthers (2009) maintains that introspection is a higher-level cognitive process in which the awareness of our mental states and the deliverance of judgments about them constitute its central function, and he *prima facie* seems to recognize that introspection is epistemically privileged from the access that we may get to other individuals’ mental states, a skeptical view survives this recognition notwithstanding.

Carruthers (2011) develops an *Interpretive Sensory-Access Theory* (ISA theory) which advocates for the acceptability of a “single mental faculty,” or a system that underlies both self-attributions of propositional attitudes and attributions to other individuals’ propositional attitudes. The class of attitudes that he tackles comprises current beliefs, thoughts, and intentions, all of which he identifies as “non-sensory accessible mental states.”

However, ISA theory contends that this single mental faculty involves a “sensory-based access” to propositional attitudes, and that this access is considered “interpretive in nature”—as opposed to its being immediate or directly accessible (pp. 1-2). That is, Carruthers argues that self-attributions of propositional attitudes “rely upon sensory cues and interpretive inferences of the sort that figure in our attributions of such [mental] states to other people” (p. 6). So access to propositional attitudes is
interpretive—that is, it obtains information about the individual’s current circumstances and current or recent behavior. It is worthwhile mentioning that Carruthers does not deny that misattribution of propositional attitudes can occur (Carruthers, 2009, p. 123; cf. 2011, pp. 19, 163, 215; see also Byrne, 2011).  

How does this mindreading faculty approximately work? According to Carruthers, this faculty gets input from several perceptual systems that broadcast both sensory and conceptual information.

First, mindreading picks up information from different sources. Sensory information, along with our perception of it, is taken from the physical environment, overt speech and behavior, proprioception, and visual and auditory imagery.

Second, conceptual information gets incorporated into the sensory representations in the course of visual and auditory processing, such as when “we don’t just see a green round object, we see an apple”, or “we don’t just hear or imagine a stream of familiar phonemes, we hear what is being said” (p. 3).

On Carruthers’ view, we use the sensory and conceptual machinery to make inferences. Our mindreading faculty can attribute propositional attitudes and can generate higher-level judgments about these attitudes. If we consider current beliefs, and if we follow ISA theory’s account, introspective access to my belief that \( p \) would depend on the aforementioned sensory input. The mindreading faculty would interpret available sensory information and thereby would attribute the belief that \( p \).

Carruthers means to say that in the introspective process we do not spontaneously attribute a belief to ourselves. Rather, we always use sensory evidence to base our attributions. That is, “whenever we take ourselves to be entertaining an occurrent thought, this should be grounded in a sensory awareness of our circumstances, our behavior, and/or sensory items held in working memory, such as a visual image or a sentence in inner speech” (p. 4).

---

35 I here bypass discussion on immediate or direct access’ accounts of introspection; also called acquaintance theories (see Ch. IV).

36 I bypass discussion on whether or not the mindreading faculty works in exactly the same way for conscious experiences.
Carruthers argues that an exhaustive survey of experimental findings research in cognitive science should be the framework that provides empirical evidence to favor ISA theory. This would include:

a. A global broadcasting architecture for sensory information, whereby sensory information, including conceptualized sensory states, is made available in a central "work-space" to specialized executive, conceptual, and affective subsystems.

b. Working memory that contains representations such as inner speech and sensory images available for global broadcast.

c. Metacognitive control of learning and reasoning, such as those involved in mindreading.

d. Disassociations between self-knowledge and other-knowledge in autism and schizophrenia.

e. Brain imaging outcomes of the systems involved in self-attributions and other attributions.

Whether or not one needs to examine all the experiments to determine whether these provide support for the ISA theory is a matter of a different discussion. To show that we not only possess a sensory-interpretive access to both our own attitudes and other individuals' attitudes, but that we also find the same method useful in achieving both self-knowledge and other-knowledge, is a large and questionable step. Since most of the cases that Carruthers presents are pathological cases, and since my project excludes these cases, I have to set them aside. Notice, however, that recent philosophical and empirical research does examine pathological cases and individuals with impaired faculties, and thus challenges ISA theory and shows that the attitudes challenge does not hold (cf. Peters, 2014a; 2014b). Let me restrict the discussion to the implications of Carruthers' view only with regards to introspection.\(^{37}\)

One of the fundamental problems with his view is the claim that the same faculty identified as a "mindreading faculty" depends on sensory input forms and is responsible both for self-attributions and for others' attributions on the basis of behavior and circumstance. In other words, sensory representations

\(^{37}\) As I mentioned, although consideration of the implications of the introspective process is relevant for subjects with pathologies and subjects with impaired conceptual or linguistic skills, I restrict this research to cases of normal adult subjects only (cf. Ch. 1). However, it is worthwhile mentioning that some cognitive psychologists are skeptical about introspective access to beliefs and claim that our access to them is theory-laden or mediated via folk psychology. Arguably, split-brain studies of patients with epilepsy show that they confabulate on beliefs by falsely associating two things. However, the patients confidently claim that that belief caused their action (Gazzaniga, 1998). Developmental studies of three-year old children show that they tend to make the same sort of error about themselves as they do about the others (Gopnik, 1993). See also Nisbett & Ross, 1980. I bypass discussion on these results.
require the same sort of mechanism of interpretation in the attribution of attitudes to others as they do in the attribution of attitudes to oneself.\textsuperscript{38}

The claim that introspection is a sensory-interpretive based mechanism and that it requires sensory-modality aspects in order to access attitudes faces a number of difficulties on different grounds (cf. Shoemaker, 1996; Rey, 2012, forthcoming).\textsuperscript{39}

Claiming that there is a single process, mechanism or common architecture, implicit in all mental attributions of propositional attitudes (whether to ourselves or to other individuals) suggests that metacognition and mindreading are considered a single process—i.e., the kind of first-person access to our propositional attitudes and the faculty to entertain representations and to attribute to other people’s attitudes something that will be underlying the same mechanism or cognitive process.\textsuperscript{40}

Some philosophers and cognitive scientists have already rejected the single process claim. They have shown that metacognition and mindreading involve independent faculties executed by different cognitive mechanisms (Nichols and Stich, 2003). Others have suggested that metacognition is prior to mindreading. So while the attribution of mental states to other individuals is dependent on introspective access to our mental states, it also involves additional processes such as inference and simulation of a certain type: “the process of re-enacting or attempt to re-enact other’s mental episodes.” That is, simulation is considered a process to read other’s minds by re-enacting the mental states of other individuals (Goldman, 2006; Shanton and Goldman, 2010).

However, Carruthers insists on “single-process” view and defends a particular version of the well-known symmetry thesis—the thesis that first-person gets access to her own mental life by using the same method she uses to access a third-person’s mental life. Nevertheless, instead of adopting symmetry across all mental conditions, Carruthers restricts his view merely to propositional attitudes. But this view

\textsuperscript{38} Notice that Carruthers (2011) prefers to use the term “mindreading” instead of “theory of mind.” So “mindreading” “is seen as neutral with respect to the realization of capacity—whether it results from theorizing, from simulation, or from the development of an innate module” (p. 2; fn. 2).
\textsuperscript{39} Harman (1990) and Tye (2000a; 2000b) advocate for an identity of sensory qualities of experiences and representational properties of experiences, whereas Block (2003) and Chalmers (2004) maintain that sensory qualities of experiences in addition to the representational properties of experience.
\textsuperscript{40} I here bypass discussion on whether the argument that posits that introspection is a single process is compelling or not. I discuss this point in Ch. V.
brings about a specific skeptical challenge about the reliability of introspection and consequently supposes that first-person access to propositional attitudes is no more reliable than third-person access.\footnote{Ryle (1949), Lyons (1986), and Dennett (1987) advocate similar views about symmetry; these views are also identified in the philosophical literature on introspection as \textit{self-other parity} views. I bypass discussion on these views.}

The consequence of accepting Carruthers’ view \textit{prima facie} undermines both \textit{privileged access}—i.e., the epistemic thesis that introspection provides a special access to one’s mental states from a first-person perspective versus the access provided from a third-person perspective, and \textit{peculiarity}—i.e., the epistemic thesis that introspection is distinct from other cognitive processes or mechanisms or that it is considered a distinct kind of awareness or a distinct way of knowing about our minds from other ways of knowing the world.

Specifically, Carruthers’ view seems to suggest the claim that first-person access to attitudes is no more reliable than third-person access or perspective, and that the same mechanism operates for both introspection—or metacognition—and mindreading. However, Carruthers’ view is not a serious challenge for introspection’s reliability.

With the aid of an attentive reading, we find that ISA theory presents two important caveats: first, that the sorts of sensory data that the mindreading employs \textit{are not always identical} between the cases of the first-person and those of the third-person. In the self-attribution of attitudes, we make use of the evidence provided by our auditory imagery or inner speech, our visual and motor imagery, and affective states. But in the attribution of attitudes to others, we do not access the third-person’s forms of imagery and affective states except insofar as these are manifested in overt speech or behavior. Second, there is generally \textit{more sensory data available} in the self-attribution of attitudes than in the attribution of attitudes to others. Yet the interpretation of attitudes available to be attributed is necessary.

The question arises as to whether these caveats can defuse the challenge or not. People might argue that it is unlikely these in particular can do so. But they do, at least partially. Although the mechanism is the same, on Carruthers’ view there seems to be a difference in both qualitative terms and in quantitative terms. Also, Carruthers perhaps recognizes a sense of agency or a self-experience as being a background of that mechanism. If so, Carruthers also holds Shoemaker’s (1996) position that the access to the contents of one’s mind is privileged: “It is of the essence of mind that each mind has a
special access to its own contents” (p. 27), and that no dispute between these positions would be at stake (see Rey, 2012; forthcoming). If that is the case, then, privilege access would be preserved, but not peculiarity.

Finally, introspection is neither an inferential process (Armstrong, 1968, p. 97; 1999, p. 115), nor does it seem to work by interpreting the sort of information mentioned above. Since introspection is considered interpretation or a sort of interpretation, Carruthers’ view would not preserve the peculiarity of introspection. Perhaps the problem lies in equating "metacognition" or introspection with "mindreading" or the inference and interpretation of other’s mental states. While a third-person inferential process is considered mindreading, it is unlikely that it is considered metacognition. Carruthers seems to distort the notion of introspection beyond recognition.

Even if we grant that the access to our minds can sometimes be based on sensory information and may sometimes require a sort of interpretation, ISA theory still rejects that a peculiar method of accessing our own propositional attitudes is at issue. However, ISA theory does not pose irrefutable conclusions or sufficient warrant such that the reliability hypothesis about attitudes can be overridden. Space prevents us from discussing criticism of ISA theory and alternative proposals to this theory (cf. Rey, 2012; forthcoming; Byrne, 2012). In Ch. V, however, I will discuss the problems of advocating for a single-faculty account of introspection and will propose to highlight the merits of a plural account of introspection.

3. The Processes Challenge

For almost forty decades, empirical research on problem-solving, subliminal perception, emotional impact, aspects of personality, and stimuli-influencing associative behavior and choices, among others, has attempted to discredit the role of introspection by reducing introspection's overall reliability. However, no research has become as dominant and paradigmatic in the experimental psychological literature concerning introspection about cognitive processes as the famous "stockings case" provided by Nisbett and Wilson (1977).

Supporters of the skeptical view about cognition have adopted the stockings case as a sound argument against introspection and have advocated for its empirical results which (supposedly) exhibit
the failures of introspection. The skeptical view claims that introspection about cognitive processes is unreliable. I have called it the processes challenge since it tackles particular targets of introspection such as the decision-making process—that is, the process of identifying and selecting alternatives based on the preferences of the individual.

Presumably, such unreliability would encompass both the introspective mechanism and the upshots or deliverances of introspection—specifically, of introspective verbal reports—even though these reports have been widely used in both experimental psychology and neurosciences. Arguably, the results of such leading and paradigmatic research into the stocking case provide empirical evidence about the unreliability of introspection about the decision-making process and its corresponding verbal reports.

I here discuss the processes challenge and claim that this challenge does not defeat the hypothesis that introspection is reliable. I argue that there are two principal problems with the stockings case. One concerns its notion of cognitive processes as targets of introspection together with its sense of "introspection"; the other concerns the unreliability of introspective verbal reports.\footnote{I am aware that an analysis of the concept of introspection and its scope entails difficulties and confusions: psychologists and philosophers sometimes take "introspection" as "retrospection", "reflection", "mere thinking", or "meditation"—e.g., Nisbett & Wilson, 1977; Schwitzgebel, 2011. Consequently, they mistakenly take questions relative to these processes as introspective questions, as I mention in §3.1.}

First, I show that the researchers of the famous stocking case fallaciously take psychological causes instead of cognitive processes themselves to be their targets of introspection; let me call them equivocal targets. It follows that the mechanisms at stake are also likely to be retrospection or reasoning instead of introspection; let me call them equivocal mechanisms (§3.1).

Second, rather than accepting the claim that reports end up being unreliable or are misinforming us on the introspection about cognitive processes, I suggest looking at the empirical evidence which proves the opposite point. Empirical evidence shows that the individual can introspectively access her cognitive processes in a reliable way. Additionally, the individual can provide accurate introspective verbal reports about her cognitive processes even if introspective reports do not exhaust the warrant or constitute the full evidence of the occurrence of the cognitive state or process (§3.2).

Let me formulate Nisbett and Wilson's argument about the processes challenge. They assert that since the individual is ignorant of the important psychological factors influencing her decisions, and since
she tends to confabulate about her decision-making process in order to express or report her behavior in an acceptable way, introspection is unreliable.

Nisbett and Wilson's famous studies aim to show that although the individual has introspective access to her mental states—i.e., her propositional attitudes and current sensations (1977, p. 255); she has no such access to her mental processes, or the higher-level cognitive processes or causes underlying her behavior and attitudes. Presumably, the results of the studies show that the individual does not have a reliable introspective access to the decision-making process. According to Nisbett and Wilson, this failure is exhibited when the participant provides introspective verbal reports about the process but those reports come out being erroneous or inaccurate.

Let me now briefly address the stocking case. It was found that individuals in a shopping mall had been ignorant of the fact that the positions of several identical pairs of nylon stockings were influencing their judgment about which pair was the best. When these individuals were asked about the reasons for their stated preferences, the individuals attributed their decisions to the quality of the stockings—e.g., whether the stockings were knit, sheer, or elastic, rather than where the stockings had been positioned in a display. When it was suggested that their choices had been influenced by the position of the stockings in the array—meaning somewhere from A in the leftmost to D in the rightmost—the individuals denied they had been so influenced.44

According to the empirical results of this case, it was found that having been asked to assess the aspects that influenced their choices, the individuals either seemed unaware of the psychological causes of their behavior, or they had verbally reported them without being aware of the intervention or manipulation of such causes. Nisbett and Wilson concluded that since the individuals' behavior had been frequently determined by implicit motives and non-conscious interpretations of the world, their introspective judgments about their decision-making processes were mere confabulations. That is, when searching for an answer to questions about their behavior, the individuals had tended to invent or to create responses, to self-attribute certain states or processes or to elaborate explanations when they did

43 For a skeptical view about introspection of attitudes, see Nisbett & Ross, 1980, pp. 200-2.
44 Similar procedures were applied for the qualities of a job applicant; subjects denied the aspects that indeed influenced their judgments and choices; see Nisbett & Bellows, 1977.
not find them. Consequently, their introspective verbal reports about cognitive processes were erroneous or inaccurate.

Even though Nisbett and Wilson’s study has been controversial, critical responses have been formulated by researchers objecting to the statement that introspection is unreliable on the basis of the stocking case’s being extreme or that the empirical procedure of their study had possessed several difficulties (Smith and Miller, 1978; Ericson and Simon, 1980; cf. Nisbett and Wilson 1977, p. 231, Wilson, 2002, p. 105). And even though contemporary philosophers and cognitive psychologists have recognized the negative effects the stocking case has meant for the philosophical and scientific investigation on the reliability of introspection, they nevertheless stress that the outcomes must be kept in a proper perspective (Goldman, 2004; 2006 p. 233; Reyes and Sackur, 2014; Marti, Bayet, and Dehaene, 2015). All told, the results of the stocking case have proved to be pervasive and (supposedly) persuasive for many skeptical advocates who have concluded that introspection is unreliable—or more specifically, that introspection about cognitive processes is unreliable (Lyons, 1986; Dennett, 1991; 2003; Johansson, et. al., 2005; Overgaard 2006; Banks and Isham, 2009; Sackur, 2009; Miller, et. al., 2010; Schwitzgebel, 2008; 2011). I cannot fully present, much less object to, any of those views here.

Beyond the theoretical risk and the philosophical provocation that this often-quoted case signifis, the skeptical view has emphasized additional difficulties. It is worthwhile mentioning at least one. Arguably, if the individual reports on the basis of the quality of the stockings—e.g., “I think this pair is the most elastic”—that it is the reason for her choice of the rightmost pair of stockings, she would be mistaken both in the cause of her choice and in self-attributing the judgment that that pair is the most elastic pair of stockings (Carruthers, 2011). Thus, not only would the mechanism of introspection and the introspective verbal reports be unreliable, but the individual’s self-attribution would also bring about its own problems.

Before proceeding with my view, I suggest that there might be an explanation for such a misleading self-attribution. Roughly, the introspective individual’s self-attribution might have been done on account of her having incorporated certain information into her decision-making model. It might be the case that she regularly applies a decision-making model while she does shopping, and therefore might tend to access the information that she has already learned on the basis of previous situations she has experienced.
For instance, the individual might have learned that the best quality products are arranged in such-and-such cases and in such-and-such a way that the products are especially accessible for the customer’s reach. However, it also might turn out that the individual’s self-attributive cause for why she chose the stockings she chose had been done on account of her having incorporated such information into her model of choice, whether or not she had been able to ascertain or to report it as such in those terms. I shall turn to examining the problems with the standard interpretation of the stocking case.

### 3.1. Equivocal Targets and Equivocal Mechanisms

At this point, an important question comes to the fore in the analysis of the process challenge: do we get access to the causes of our cognitive processes via introspection? While it is very likely that we can achieve introspective access to our cognitive processes such as decision-making, to suppose nonetheless that we can get introspective access to the causes of those processes is a large and questionable step.

Claiming that we have access to our decision-making process is different than claiming that we have access to the causes of our decision-making. While the former requires an introspective focus on our currently ongoing process, the latter requires further self-probing; it extends beyond our currently ongoing process and focuses on supplementary information.\(^\text{45}\)

A simple answer to the previous question may be: if we do not have introspective access to the causes of our cognitive processes, it is possible that those causes either remain unconscious or could be accessed via a different mechanism or through the intervention of additional mechanisms and cognitive processes in tandem.

It is plausible that some psychological causes of our decision-making process or motives for acting remain unconscious—e.g., the causes may be controlled by social pressure, or may be unconsciously displaced or substituted for another process. Alternatively, other defense mechanisms such as repression or denial may be in play when trying to access those causes.

---

\(^{45}\) The advocate of the skeptical view, however, might object that if one were introspectively aware of a process such as an algorithm by which one decides, then one would be aware of the information used in the process—e.g., the position of the stockings in the array.
In fact, if we do a close reading of the stocking case, we might find that Nisbett and Wilson themselves could be content with this suggestion or might even reinforce it: “Individuals are sometimes (a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response” (Nisbett and Wilson, 1977, p. 231).

Moreover, the causes of our cognitive processes or underlying mechanisms are not prima facie available through introspection. Most importantly, however, the inability to determine psychological causes via introspection would not constitute a failure of introspection, but rather a restriction of introspection’s access or a limit of introspection (cf. Ch. II, §4.2.). It is unlikely that the putative psychological causes can be conceived of as targets of introspection; hitherto, according to a general agreement concerning the contemporary literature of introspection, propositional attitudes, conscious mental states, concurrent actions, and mental processes have already been established as the typical targets of introspection. So based on the stockings case, it is reasonable to identify the psychological causes of our decision-making process as (what I call) equivocal targets.

It is reasonable instead to suggest that the psychological causes of the cognitive processes and what brings us those processes require further probing, since these extend beyond our current mental states, immediate experiences or ongoing processes, and focus on supplementary information. Thus, in order to get access to these causes, a rational reconstruction would be required or an interference of mechanisms, or those higher-level cognitive abilities capable of discerning the abstract relations between causes and processes, would also be needed. Introspective training and practicing would be needed, too. But it is not my intent to cover this point here (cf. Ch. II).

In a nutshell, being aware of our cognitive processes is different from being aware of the causes of our cognitive processes. Again, while the former sense of awareness refers to the typical targets of introspection, the latter refers to targets of rational reconstruction or reasoning, and/or the intervention of additional mechanisms.

So instead of seeking the cause of the decision-making process, what is at stake here is the kind of access the individual has to self-probe her ongoing process. This might include how her decision-making about a pair of stockings appears to her from a first-person perspective, provided that certain
conditions obtain; how she forms corresponding judgments about her cognitive process based on an awareness of that mere occurrence; and how she offers introspective verbal reports of the occurring cognitive process *ceteris paribus.*

To illustrate: suppose you generate an introspective judgment on your decision-making process about what pair of stockings to choose. You might have in mind to use the stockings in your dance class. And so you might entertain a corresponding mental imagery: the execution of a series of bodily movements you make while wearing them. Alternatively, you might wish to get the stockings simply because winter is arriving. And so you might entertain a corresponding mental imagery: wearing layers to avoid catching a cold. Although either or both or neither of these might be count worthy towards the causes of your decision-making process, you are still able to access your occurring process introspectively and, then, verbally report on it—even if you do not tackle its causes or even if you do not find out the underlying cause of such a choice. It might be the case that such a cause is the mere desire of alleviating a varicose-vein pain.

Now, insofar as you generate an introspective judgment of the cognitive process at issue—e.g., choosing one pair of stockings after having selected a specific belief among several alternatives, then carrying out the choice that prompted the action—you can rely on your introspection of that particular process. Alternatively, you can consider introspective access as a basic evidential source of your occurring process. I also admit that insofar as a new piece of evidence or relevant information on your decision-making process could be at stake, the possibility exists of there being still some further revision in order for you to determine the *range* of reliability of your introspective access.

It is important to be clear that an introspection of the decision-making process entails the introspective judgment and/or the corresponding verbal report of the present choice from a first-person perspective, irrespective of any association, influence of behavior, or further evaluation. But Nisbett and Wilson seem to neglect that the *causes* of our cognitive processes are not targets of introspection. Moreover, as it is unclear what the scope of introspection is, it is not obvious on their views what the boundary conditions of introspection would be. Since the processes challenge appeals to *equivocal targets* rather than to the decision-making process *itself,* this challenge does not defeat the hypothesis that introspection is reliable.
Notwithstanding these points, some philosophers might object and say that Nisbett and Wilson recognize that the stockings case is actually a case of *retrospection*. But if this were the case, it would be necessary both that they clean up the negative effects that their empirical research has brought about on introspection’s reliability, and propose a new direction in the philosophical and scientific study of both introspection and *retrospection*, defining or redefining these concepts in the first place.\(^{46}\)

Still other philosophers might object that Nisbett and Wilson’s research presents a number of additional experiments that emphasize the *failures* of introspection, all of which are exhibited in the individual’s introspective verbal reports. For example, Nisbett and Wilson claim that the influential experiment of Maier (1931)—i.e., the two-rope case—“suggests that even when individuals are thoroughly cognizant of the existence of the relevant stimuli, and of their responses, they may be unable to report accurately about the influence of the stimuli on the responses” (1977, p. 242). Nevertheless, *pace* the advocates of this research, the two-rope case instead rests on *problem-solving*—that is, a target of a different cognitive mechanism—as well as on other cases that allude to memory and aspects of personality which rest on different mechanisms and/or a combination of them.

Let me sketch out the two-rope case. Two ropes were hung from the ceiling of Maier’s lab after which experimental individuals were asked to tie the two ends of them together. However, the ropes were placed far apart, so the individuals could not reach one rope while holding on to the other. The individuals were given the opportunity to provide solutions to this problem by using a stick, a pair of pliers or weights, and a chair. When asked to provide new solutions, however, the individuals were initially incapable of doing so. But, when Maier accidentally gave them a clue when he set one of the ropes in motion, the individuals responded within a minute by first tying a weight to the end of one rope and setting it in motion, then waiting for the rope to swing close enough for them to grasp it. Yet even though the individuals were eventually able to provide a solution, they were unaware of the precise cause of their behavior—i.e., that it was Maier’s clue which had brought about the solution. And so they failed in their reporting of it. That is to say, they merely confabulated (Maier, 1931).

\(^{46}\) For problems on the concept of introspection and a proposal on species of introspection, see Prinz, 2004. For discussion on kinds of access or different methods to report, see Overgaard & Sandberg, 2012.
Notwithstanding all this, if we set aside for a moment the alleged failures in the act of verbally reporting, it is doubtful that consideration on the Maier’s famous experiment would count as a case of introspection. Again, a problem-solving process such as the two-rope case, in which the individual is required to figure out what must be done to achieve a particular task, involves an additional effort such as an understanding of the process in a logical way and the drawing of inferences and conclusions. So it is reasonable to suggest that this might be the target of a cognitive process such as reasoning.

Reasoning differs from introspecting in several ways. When we are introspecting our cognitive states or processes, judgments may simply come to us, one after another, in a somewhat random way. On the other hand, when we are reasoning, we make inferences or link thoughts together in a way that favors sequencing each judgment so that it will provide support for another judgment or thought. The inferential process entails a particular relationship between different thoughts. For instance, to solve the aforementioned problem where the experimental individual infers setting the ropes in motion from thinking about motion—after accidentally seeing a rope in motion—she will be moving from that specific thought in order to solve the problem in question. This makes it seem reasonable to believe that there could be a possible solution, even if the exact cause of the process remains unconscious or seems to be unconscious.

Hence, instead of alluding to introspection per se, other cognitive processes such as inference or reasoning seem to be at stake in the inquiry of the psychological causes of our cognitive processes. Since the two-rope case appeals to other mechanisms rather than introspection, it would be reasonable to suggest there is (what I call) an equivocal mechanism at play.

In support of this suggestion, updated empirical evidence shows errors in the empirical procedure of Nisbett and Wilson’s research. Even if we do not elaborate at this point, contemporary scientists claim that “the cognitive processes that Nisbett and Wilson target are complex, high-level forms of reasoning” (Reyes and Sackur, 2014, p. 213). So not only would introspective access not be playing a role here; the same issue would be applying to questions and provision of verbal reports. When asked about the causes of her cognitive processes, the experimental individual would need to deploy either the intervention of additional mechanisms or the use of additional cognitive and psychological capacities.
To offer a response to such a question about her cognitive processes, the experimental individual would be required to draw conclusions from certain facts or evidence, and then sort out an explanation of the cause or offer an argument of the cause at stake. Recall that not all questions can be considered introspective questions (cf. Ch. I). Asking for the causes of our processes or what we can identify as “high-level [cognitive] questions should not be used [on these kind of experiments] because reports will be tainted by confabulations” (Reyes and Sackur, 2014, p. 217).

The lack of a cogent defense of the skeptical view is particularly troubling in the light of recent empirical evidence showing that introspection of cognitive processes is reliable—pace the skeptical view. Experimental individuals can indeed get accurate information about their cognitive processes during visual search tasks, provided that the target of introspection is elementary and short, that there is guidance which promotes attention in the target to search, and that introspection is performed systematically and immediately after the task. Individuals can then focus on the process of interest and thus provide accurate introspective verbal reports (Reyes and Sackur, 2014).

Since the processes challenge appeals to equivocal mechanisms rather than to introspection, this challenge does not defeat the hypothesis that introspection is reliable.

Before closing this section, some words on recognition of error might be helpful. While the research was meant to deny that individuals have introspective access to cognitive processes or the causes of their behavior, Wilson himself (2002) recognizes that their research (1977) contained problems. So the claim that people do not have introspective access to their decision-making process was incorrect.\(^{47}\) Says Wilson: “Perhaps the most controversial part of our article was the claim that people have limited introspective access to their mental processes. Any sentient human being knows that an extreme version of this argument is false. The fact that people make errors about the causes of their responses does not mean that their inner worlds are a black box…” (p. 105).

Wilson argues that he and Nisbett denied introspective access only to the causes of cognitive processes or higher-level processes, not to the processes themselves.

\(^{47}\) Contemporary supporters of this research do not seem to be aware of this fact yet—e.g., Schwitzgebel, 2008; 2011; Carruthers, 2011.
It is true that people have privileged access to a great deal of information about themselves. The real action in the mind is mental processing that produces feelings, judgments, and behaviors. Although we often have access to the results of these processes [...] we don't have access to the mental processes that produced them [i.e., their causes]. I don't really know, for example, why [...] a particular memory came to mind, just as the participants in the panty-hose study did not know exactly why the preferred pair of D over A.” [...] The Nisbett and Wilson argument can be reworked as follows:

- Many human judgments, emotions, thoughts, and behaviors are produced by the adaptive unconscious.
- Because people do not have conscious access to the adaptive unconscious, their conscious selves confabulate reasons for why they responded the way they did [...] (Wilson, 2002, pp. 105-6).

Notice that Wilson explicitly claims “we do not have access to the mental processes that produced” those processes. Further reading shows that they might have tackled equivocal targets and, thus, equivocal mechanisms as well. It is also likely that they might have conceived attitudes and conscious experiences to be processes instead of states. But this is a matter of a different discussion that I do not tackle here.

In any case, it is worthwhile mentioning that “[p]erhaps the most radical part of [the] Nisbett and Wilson argument is that despite the vast amount of information people have, their explanations about the causes of their responses are no more accurate than the explanations of a complete stranger who lives in the same culture. Even though we do have a lot of more information about ourselves than a stranger does, however, this information may not always lead to accurate inferences about the causes of our responses” (Wilson, 2002, p. 109, my emphasis; see also Wilson, 2003; Goldman, 2004, pp. 233-5).

For present purposes, it is convenient to emphasize an additional detail. Nisbett and Wilson (1977) concede: “We wish to acknowledge that there are methodological and interpretive problems with some of the individual studies... We also wish to acknowledge that the studies do not suffice to show that people could never be accurate about the processes involved... What the studies do indicate is that such
introspective access as may exist is not sufficient to produce accurate reports about the role of critical stimuli in response to questions asked a few minutes or seconds after the stimuli have been processed and a response produced” (p. 246).48

To sum up, the stocking case has been the paradigmatic example that underpins the alleged unreliability of introspection for cognitive processes and their corresponding verbal reports. But taking the results of this research at face value engenders serious problems for the philosophical and scientific study of introspection. As we have seen, Nisbett and Wilson do not deny that we can introspectively access our cognitive processes. But their argument is wrong given that we can introspectively access the causes of our processes, and we can consider the causes of processes as processes themselves without implying that different mechanisms or interference of additional higher-order mechanisms may be at stake.

Whether the individual can or cannot introspectively access the decision-making process would require a new experiment that precisely indicates what the introspecting decision-making process consists of. So far, the alleged unreliability of the introspective access to the cognitive process has led to mistakenly distrust introspection. Therefore, it might be time to make amends.

4. Introspective Verbal Reports about Cognitive States and Processes

In this section I discuss recent empirical results that emphasize the role of introspective verbal reports—which are also called upshots or deliverances of introspection—as the primary sources of evidence favoring introspection’s reliability both of cognitive states, such as propositional attitudes like beliefs, and cognitive processes, such as the decision-making process.

Although introspection does not necessarily result in verbally reporting our mental states or processes aloud—we may generate judgments about them and these can be either reported or remain in inner speech—it is important to distinguish between introspective judgments, introspective verbal reports, and the introspective process itself. Judgments and/or reports may count as steps or parts of the

---

48 Reyes & Sackur, 2014 provide a number of solutions to these problems. Space prevents us from discussing them.
introspective process, but they are not reducible to it. So it is untenable to claim, as the skeptical view does, that because introspective verbal reports are misleading, introspection itself is unreliable.

On the contrary, empirical psychologists take advantage of those introspective verbal reports or narrative modes which describe first-person perspectives by presenting either the equivalent of one’s stream of consciousness, or the self-probing process typically connected to inner speech that shows the self-attribute and judgments of cognitive states, processes, and so forth.

Cognitive states and processes are targets of introspection in virtue of which one introspectively provides verbal reports to the effect that one can access particular states and processes, as well as their character and contents. These can include offering distinctions of propositional attitudes according to types and contents, for example, and justification of the introspective judgments about one’s ongoing mental life.

Recall that the process challenges comprise the alleged failures of introspection about cognitive process and the inaccuracy of introspective verbal reports. In previous sections, we have cleared up some skeptical doubts. Nevertheless, it still is necessary to describe how the story can be modeled in the framework of an account of introspective verbal reports.

Nisbett and Wilson (1977) suggest that even if we get some introspective access, it does not suffice to provide regularly and systematically reliable verbal reports, given that the experimental individual either bases her introspective verbal reports on a priori theories about the causal connection between stimulus and response, or ends up ignoring the trace of her explanations and confabulates on her states or processes.

Disagreeing with this perspective, I again claim that process challenges do not defeat the thesis that introspection is reliable. My strategy here is to offer recent empirical evidence to show that the individual is able not only to access introspectively her cognitive states and processes in reliable terms, but that she is also able to provide accurate introspective verbal reports about them—again, even if the reports do not exhaust the warrant or constitute absolute evidence of the process in question.

Certainly, in order to report the causes of the decision-making process, the introspective individual may engage in theorizing beyond what is available in her states or processes. Empirical research has detected biases that arise in virtue of what the individual considers to be the cause of her
choice (Lutz and Thompson, 2003; Ericsson and Simon, 1980). However, this theorizing may be grounded on a decision-making model that entails the incorporation of relevant information on account of past experiences without thereby suggesting a problem of reportability, strictly speaking.

To elaborate: even if the individual were ignorant of the causes of her mental process—e.g., the way in which aspects of her personality influence her behavior, the reasons for being attracted to someone, why noise affects her enjoyment, and so forth, as Nisbett and Wilson observe, and even if the individual were able to make judgments about how plausible it is that a stimulus has influenced her response by adding to the description of the occurring process to make sense of it—confabulation of explanations or reasons does not suggest a problem to introspection. The individual verbally reports from a first-person perspective and according to how the process appears to her. Also, her introspective reports are based on descriptions of her own current, ongoing or recently past mental process; not on explanations of other aspects such as their psychological causes or reasons.

Moreover, recent empirical research shows that if the experimental procedure is simple—thus, simplifying tasks, avoiding noises or distractors, reducing response time patters, and guiding the experimental individual to attend to specific targets—it is possible to reduce or to avoid confabulation (Reyes and Sackur, 2014; Reyes forthcoming, and personal communication).

Besides, empirical research suggests that confabulation has been typically demonstrated in individuals with certain pathologies—patients who suffer amnesia, or act out posthypnotic or split brains, and individuals with impaired conceptual or linguistic skills (cf. Gazzaniga and Le Doux, 1978; Wilson, 2002). However, even if normal adult individuals add to their descriptions, their introspective report might be open to revision to determine a range of reliability, insofar as a new piece of evidence or further relevant information is at stake.

To be clear: a misleading upshot or deliverance of introspection does not count as a failure of introspection, since it may be the case that certain evidence only defeats an instance or a part of the deliverance. Also, a misleading upshot or deliverance of inspection does not defeat all upshots or deliverances of introspection, nor does it deprive the introspective process from reliability. Thus, confabulation does not suppose a real problem to introspection. On the other hand, as I previously
mentioned, upshots or deliverances of introspection can be open to revision to determine the range of reliability insofar as a new piece of evidence or further relevant information is at stake (cf. Ch. I-II).

Some philosophers claim that introspective verbal reports can be considered “provisional warrants” of introspective access of the occurring cognitive process in question (Goldman, 2004). So it is reasonable to hold that when the individual provides an introspective verbal report on her decision-making process about a specific pair of stockings, for example, that report would account for a particular deliverance of the process that is occurring in her mind—i.e., a *prima facie* warrant of the occurrence of such a process—unless it were overridden by additional evidence.

Similarly, if the experimental individual introspectively reports on that process to an experimenter, the experimenter would be *prima facie* warranted in believing the individual’s report even without having monitored her brain activity or having tracked a record on the cognitive-processing. Nonetheless, if additional evidence concerning the facts of the individual come to the fore, it would be necessary to assess whether the report does or does not merit even provisional reliability. If the introspective individual can generate judgments about her current process and provide corresponding verbal reports based on her awareness of that occurrence, introspection would be considered the “evidential source that confers *prima facie* warrant” of those judgments and reports—even though a method could never be “perfectly reliable” (Goldman, 2004, pp. 3-4).

It is reasonable to suggest, then, that introspection’s reliability is given as a function of the strength or consistency of its deliverances, although it can never be reduced to it. So far, the stocking case has not shown that the deliverances of the occurring cognitive processes conflict in such a way that all deliverances of introspection or its reliability could be reduced or weakened.

Nevertheless, in support of the stocking case, some empirical psychologists have claimed that since introspection of cognitive processes entails a certain sort of “re-representing” targets and its contents, it might lose some information or become distorted when the individual generates judgments and/or describes her process in the form of verbal reports. They have referred to this problem as “verbal overshadowing” (Schooler and Engstler-Schooler, 1990; Schooler, 2002). So if the introspective verbal reports are inaccurate or misleading, the reliability of introspection might be threatened. However, empirical evidence can provide the *conditions* to optimize introspection’s reliability via the accuracy of
introspective reports. Wilson himself (2003) observes that there is evidence for the reliability of introspective reports in several areas of empirical psychology (p. 131), and Nisbett and Wilson's (1977) research identifies circumstances which promote error in verbal reports, and predicts when the subject will be correct in her verbal reports as well (pp. 251-3).

Cognitive scientists advocate for controlled conditions when using introspective reports; they take reports prima facie as direct sources of evidence of individuals' mental lives and as insight into the operation of target processes themselves (Jack and Roepstorff, 2004). Certainly, contrary to behavior which can be easily categorized as being right or wrong, successful or unsuccessful, "reports can't be so easily verified." However, in order to determine whether reports are accurate, these need to be revised if grounded on both independent evidence and evaluation of reliability, such as on the "individual's categories and conceptual frameworks for understanding the world," the "relation of trust and expectations between experimenter and experimental individual," and, insofar as possible, the provision of literal, honest, and undistorted descriptions ((Jack and Roepstorff, 2004, p. xix).

Empirical evidence shows that reliable deliverances obtain when experimental individuals report on their cognitive processes through "think-aloud protocols." That is, experimental individuals provided with specific directions are asked to report the thoughts that cross their mind as they engage in a task, while the experimenter registers the conditions under which the individuals report and offers reliable data that correspond to behavior without affecting performance. More specifically, individuals' think-aloud-protocols are found reliable when they are done concurrently without the individuals' explanation of the processes underlying their thoughts. Researchers have been able to distinguish reports that are prone to be unreliable or "reactive"—that is, when the retrospective reports or those reports that require that individuals make inferences about why they behave in a particular way (Ericsson and Simon, 1980; 1993; Ericsson, 2003).

In a nutshell, Ericsson and Simon found that when researchers tried to get more information by asking experimental individuals why they had responded in a certain way—"Why do you like him?" "How did you solve this problem?" "Why did you take that job?" as Nisbett and Wilson (1977, p. 231) have done—no expectation of getting reliable reports could be fulfilled. So the think-aloud protocol, also called the protocol-analysis method, not only avoids reports' tendency to mislead, but it also offers experimental
and theoretical accounts to identify different types of introspective verbal reports that provide the greatest possible range of reliability.

In fact, recent empirical research has tested Ericsson and Simon’s (1980) work and has confirmed introspection’s reliability via reports; experimental individuals can provide partial but detailed and reliable introspective reports about their cognitive processes during task execution. In exploring consistencies and differences between introspective reports and actual behavior within a single trial—eye movements and individual’s beliefs of where their eyes moved having been recorded—researchers have found that the comparison of reported versus real eye movements shows that experimental individuals successfully report a subset of their eye moments. Even though introspective reports might provide limited accuracy, they provide information that matches individuals’ objective behavior without distortion (Marti, Bayet, Dehaene, 2015, p. 13).

Contrary to Nisbett and Wilson’s findings, contemporary researchers have offered empirical evidence about the reliability of introspecting cognitive processes and corresponding verbal reports. That is, we can introspectively access our mental processes in reliable terms by making use of controlled, rigorous, and systematic conditions. Current experiments show that experimental individuals, when they are clearly guided, can be both introspectively aware of those processes and become better aware of their processes. This leads us back to the relevance of learning, training and practice (Ch. I, §1; Ch. II, §3). That is, experimental individuals can introspectively access their processes through specific mental acts, such as prompts and questions, and can provide reliable verbal reports on them. To avoid explanation and rationalization, which tend to substitute for the actual target of introspection, the experimenter conducts an “elicitation interview” to guide the individuals to introspect and report on their occurring process (Petitmengin, et. al., 2013).

In order to undertake an elicitation interview, the first step, according to the researchers, is to distinguish between decision-making criteria and the decision-making process. Scientists have shown that we are generally aware of our decision-making criteria, but we are frequently not aware of our decision-making process. Nor is it easy to be introspectively aware of our decision-making process. To illustrate: let us say that we can be aware of the criteria we are using to choose one among several outfits when get dressed in the morning. Still, we can be frequently unaware of the cognitive operations we are
using as we make our choice. For example, we might not be aware of what the inventory of our day’s activities is going to be, or what a visualization of ourselves in a meeting using that particular outfit we chose would look like, or what a mental image of our feet after walking with new shoes for some hours would be, etc. (Petitmengin, et. al., 2013). And yet the researchers are not claiming that we do not introspectively access decision-making processes; rather, they are trying to point out that simplicity and regularity are not features of introspection of decision-making processes.

Moreover, empirical evidence shows that introspection of the mental processes may require a “particular expertise.” The elicitation interview is premised on an individual’s acquisition of that expertise. When experimenters guide individuals to choose a particular occurrence of their decision-making process to be described, instead of asking: “why did you choose this picture?”—which is what the protocol of Johansson et. al., (2006) does in order to garner explanations or generalities about the individuals’ processes—an elicitation interview would carry out the direction of specific aspects and precise tasks.

Most importantly, recent results show a significant difference concerning two scenarios—one involves counting with the help of such “elicitation interview” and one involves counting without it. For example, by reproducing the “choice blindness” paradigm (Johansson et. al., 2006; based on the research of Nisbett and Wilson, 1977), the experimental participant might be presented with a pair of photos of women and asked to choose her preference. After six trials, the experimental participant is handed her chosen photo and asked to explain her choice. In three of those trials she is “secretly handed” the non-chosen photo; in three of those trials she is not so handed it. Interestingly, participants were able to detect the manipulation of their choice in 80% of the cases with the help of elicitation interview, whereas participants detected the manipulation in only 33% without the help of elicitation interview (Petitmengin et. al. 2013; Froese, 2013).

These results contribute to determining the circumstances in which the individual can succeed to a considerable extent in the reliability of her introspections. They also make possible a way to describe the cognitive-processing at stake. Results suggest that an introspection of cognitive processes can be rich and reliable thanks to scientists having undertaken several experiments of their contents where they were able to detect regularities in a set of choice strategies.
5. Conclusion

To sum up, several problems have come to the fore in the contemporary philosophical and scientific debate on introspection of propositional attitudes. I have discussed those problems and offered specific responses to the alleged failures of introspection. I have shown that the attitudes challenge does not defeat the hypothesis that introspection is reliable since the targets of introspection are occurrent beliefs instead of standing attitudes.

Further, I have shown that Carruthers’s ISA view is unpersuasive and does not defeat the hypothesis that introspection is reliable, since his view alludes to interpretation instead of introspection; introspection holds an asymmetry between first-person and third-person perspectives and Carruthers’ view does not.

Most importantly, I have shown that the process challenge does not defeat the hypothesis that introspection is reliable, since psychological causes of our cognitive processes can be conceived of as targets of retrospection, as opposed to targets of introspection.

I have offered arguments and empirical evidence to show that the process that the paradigmatic and dominant research by Nisbett and Wilson attacks is very likely to be a process of reasoning, and not of introspection. I have suggested there are at least two difficulties with the stocking case: equivocal targets and equivocal mechanisms. I have also shown that there are errors in the experimental procedure of the research which do not endorse the process challenge at all. On the contrary, the recognition of error by Wilson should be reconsidered before a final balance concerning introspection’s reliability is determined.

Additionally, I have discussed the role of introspective verbal reports and have maintained that they preserve a prima facie evidential warrant or status for the introspection of our occurring states and processes. Insofar as a new piece of evidence or further relevant information is at stake, however, this warrant would be open to revision to determine the range of reliability.

Finally, I have tackled recent scientific studies which show important aspects of the reliability of introspection about cognitive processes. Through the “protocol-analysis method” or “elicitation interview,” for example, we can confirm both that our introspective reports are accurate, and that these correspond to the cognitive process in play.
Before proceeding onward to examine leading philosophical accounts of introspection (Ch. IV), it is important to recall that skeptical doubts about the reliability of introspection have been cleared up in previous chapters (Ch. I-III). I have defended the view that introspection holds qualified forms of privilege access; it provides a reliable access from a first-person perspective versus the access of one’s mental states provided from a third-person perspective.

So far, I have argued that the reliability of introspection is given as a function of different epistemic conditions of introspective access. This includes both qualified forms of privilege access—viz. judging on the base of/in consideration of context, no generalization of introspective target states, authority of error, possession of right cognitive capacities—and special conditions of introspective access—viz. paying attention, being cautious, applying corresponding conceptualization, being able to offer a justification of our currently ongoing and recently past mental states, and consideration of particular situations.

The background for addressing skeptical challenges to introspection and the boundary conditions in which introspection is successfully reliable have been set up (Ch. I), and the alleged skeptical challenges to introspection’s reliability—the across-the-board challenge and the restrictive challenges—have been cleared up (Ch. II and Ch. III). Now, let us turn to examining the philosophical accounts on the nature of introspection.
IV. Anti-Reductionism and the Nature of Introspection

What is introspection? Several contemporary philosophical models of introspection have been offered, yet each faces a number of difficulties in providing an explanation of the exact nature of introspection. I here contrast the inner-sense model that argues for a causal awareness with the acquaintance model that argues for a direct awareness. After critically examining the inner-sense and the acquaintance models, I claim that these two views are complementary and not mutually exclusive, and that both perspectives, conceived of as two functional levels of introspection, can be reconciled and actually broaden the notion of introspection. I then propose a distinction between (what I call) stimuli-induced introspection—that is, a passive or receptive process whereby some specific mental states cause introspection—and (what I call) self-triggered introspection—that is, an active or selective process whereby the individual’s own interest and volition initiates introspection. I argue that that distinction may eliminate the false dichotomy which claims that only one of those types of accesses, either the causal one or the direct one, is conducive to introspection or is defined as introspection.

1. Introduction and Preliminary Considerations

Any philosophical model that aims to provide an account of the nature of introspection must explain which conditions are necessary and sufficient for introspection. From that move, it is possible to determine the specific nature of this cognitive process and the psychological mechanisms that underlie introspection.

Many contemporary philosophical theories of introspection have been presented, yet no consensus has been reached about the nature of this cognitive process. Introspection is said to be either (A) a kind of perception defined as an inner-sense or a mechanism of self-detection operating as an internal scanning or monitoring of our mental life—henceforth, the inner-sense model (e.g., Armstrong, 1968/1993; Lycan, 1996)—or (B) a kind of knowledge by acquaintance operating as direct awareness which can provide justified, non-inferential, judgments of our mental life—henceforth, the acquaintance model (e.g., Gertler, 2011, 2012; Chalmers, 1996, 1999, 2003).

Although the basic conditions of introspection—such as introspection is (a) directed at one’s own mind (first-person), (b) is about psychological states and processes—namely, mental entities and

---

49 Space prevents us from discussing other accounts that might be considered theories of introspection: a “higher-order state” (Gennaro, 2012), a “higher-order perception” (Lycan 1996), a “third-order thought” targeting a current mental state (Rosenthal, 2005), or a “meta-awareness” distinct from merely having an experience (Jack & Shallice, 2001; Schooler & Schreiber, 2004). My discussion is intended to be neutral concerning these accounts. Theories that have normative implications are not relevant here, such as “rationalist” accounts that involve rational conditions, responsibility, or rational agency for attitudes via practical reasoning (Burge, 1996), deliberation or theoretical reasoning (Moran, 2001). Neither are “self-shaping,” “self-fulfillment,” or “containment” accounts relevant which claim that introspective judgments shape, create, or contain the target mental state, or argue that judgments involve pre-existing states, current states, or immediate future states (Dennett, 1991; 1987; Hogan & Kriegel, 2007; Chalmers, 2003, 2002; Shoemaker 1994).
properties, as opposed to non-mental entities and properties (mental), and (c) is about one’s current, ongoing, and recently-past mental life (occurring states and processes)—ground the analysis of introspection’s nature, current debates focus on whether these conditions are sufficient or not.

Advocates of the most prominent models of introspection—that is, those of the inner-sense and acquaintance kind—consider the previously-mentioned conditions insufficient. These advocates believe that in order to define a cognitive process as introspection, additional conditions have to be met. Respectively, the process has to be either causally connected to the target mental state, thereby involving a certain mediate awareness (henceforth, causal), or non-causal and hence directed to the target mental state, thereby involving an immediate awareness (henceforth, direct).

Advocates of the inner-sense and acquaintance models, then, base their accounts on ostensibly opposing conditions: they claim that introspective awareness is either causal or direct. Notice that the conditions are given as a function of the specific relation that introspection qua cognitive process bears with its objects: its target mental states.\(^\text{50}\)

Although the contributions of the inner-sense and the acquaintance models have been significant in shedding light on introspection, they face a number of theoretical difficulties in providing an explanation of the exact nature of introspection. After critically examining the pros and cons of the inner-sense and the acquaintance models, I claim that those views are complementary and not mutually exclusive, and that both perspectives, conceived of as two functional levels, can be reconciled and can even broaden the notion of introspection.

I then propose a distinction between stimuli-induced introspection—a receptive process whereby some specific mental states cause introspection—and self-triggered introspection—a selective process whereby the individual’s own interest and volition initiates introspection. I argue that this conceptual distinction, in terms of modes of introspective access, may eliminate the false dichotomy which claims that only one of the mentioned conditions, either the causal one or the direct one, is conducive to introspection or is defined as introspection.

The rest of this chapter has five main sections. In order to set forth the most important aspects for the debate on the nature of introspection, §2 points out some of the basics of the inner-sense model and

\(^{50}\) I refer to qualitative states, aka, conscious states, phenomenal consciousness, or non-cognitive states.
§3 points out some of the basics of the acquaintance model (readers familiar with these views might want to omit §§2-3). Next, §4 clears up misconceptions of both “inner-sense” and “direct,” and spells out the crucial problems of these models. §5 suggests that two functional levels of introspection may be at stake. Finally, §6 argues for my own proposal to broaden the notion of introspection.

2. The Inner-Sense Model of Introspection

The inner-sense model of introspection (ISMI) defines introspection as a kind of perception, an inner-sense or a mechanism of self-detection operating either as self-scanning or an internal monitoring of our mental life. This model has its antecedents in some traditional views which claim that introspection is an “internal sense” whereby we observe the operations of our mind in the same way we perceive physical/external objects or objects of the environment (Locke 1689/1975).

While some philosophers have endorsed this view by claiming that introspection is “‘perception’ according to a Lockean ‘inner sense’—i.e., by the act of one’s focusing one’s attention on the internal character of one’s experience itself” (Lycan 1996, p. 4),—others claim that introspection is strictly speaking an inner sense (Kant 1781/1929), or a self-detection mechanism for monitoring experiences whereby we “become aware of current happenings in our own mind” (Armstrong, 1968/1993, p. 95/323).

For a concrete view of this position, Armstrong (1993) serves as a representative for ISMI. He construes introspection as an inner-sense or perceptual-like mechanism: “a mental event having as its (intentional) object other mental happenings that form part of the same mind.” Moreover, he defines introspection precisely as “a self-scanning process in the brain’ encompassing “a mere flow of information,” which then identifies “mental states with material [or physical] states of the brain” (pp. 323-4, 326 my emphases).

Armstrong maintains the basic conditions of introspection that we have encountered in the introduction of this chapter and in previous chapters (centered on the first-person, the mental, and the occurring), however, he appeals to the causal condition. That is, introspection “is confined to our own minds … [W]hen I acquire by introspection the information that… I am sad now… this information [is] about… my behavior-producing or potentially behavior-producing states.” Armstrong notes that introspection provides “information… about the current state of our mind” (pp. 325-6) and “is the acquiring
of information (or misinformation) about our own current mental states [which]... qua mental states [make]... the person apt in their various ways for the production of certain sorts of physical behavior" (p. 333).

For instance, a pain such as a headache indicates that the suitable mental state which has been caused by a particular stimulus consequently causes a specific behavior. Hence, for the advocate of ISMI, introspection is a process of a system that is determined by its causal relations to other states and/or processes. Additionally, introspection grasps mental states as potential causes of behavior, since mental states can externally exhibit or produce physical behavior—e.g., an expression of enjoyment as a result of a tasty taco or a manifestation of an intense craving for grasshoppers.

In order to account for a causal condition of introspection, Armstrong first distinguishes between a current target mental state and the introspective awareness of that mental state. Building on a parallel between perception and introspection, Armstrong then defines introspection as a causal process which involves scanning or monitoring the ongoing mental state. Armstrong claims that “it is an essential mark of veridical perception that the situation that is perceived is the cause of the perception.” Likewise, he asserts, “where it is veridical, the mental state of affairs that we are aware of brings about the [introspective] awareness of it” (p. 329). Notice that a state is veridical simply when it accurately represents its content.

It is worthwhile emphasizing that Armstrong claims that there is no introspective awareness of the causal relation itself, nor any detection of “the intrinsic features of the mental” which he considers “physical” (p. 324). In other words, according to him, neither the causal connection between the introspective awareness and its mental states, nor the physical features of the ongoing state, is accessible through introspection. For instance, introspecting a qualitative state such as a pain “does not provide us any special insight into the nature of pain, because even if we can determine, by introspection, that pains cause a strong desire that they should cease, it cannot capture its intrinsic features (physical features) or access its basic nature” (p. 314). Consequently, certain information that the mental states bear is not available via introspection. This information, however, can cash out as physical properties of the mental states or “sub-personal” states, borrowing Dennett’s term (1969).
To summarize, ISMI construes introspection as a faculty of inner-sense or a “quasi-perceptual” process that “features inputs—that is, events or properties to which the process is causally sensitive—and outputs—that is, representations generated in response to those inputs” (Goldman, 2006, pp. 225, 246). The inner-sense faculty simply operates as the monitoring or scanning of our own mental life. Finally, a monitored or a scanned mental state, such as a visual experience, is considered the input of introspection; whereas a self-attribution and its corresponding introspective judgment of that experience is considered the output of introspection, thus giving rise to a causal relation between the introspected target mental state and the process of introspective awareness.  

3. The Acquaintance Model of Introspection

Like ISMI, the acquaintance model of introspection (AMI) maintains the basic conditions of introspection: first-person, mental, and occurring. But contrary to ISMI, AMI claims that the relation that introspection bears with its target mental states is direct rather than causal. For it holds the direct as a necessary and, together with the three basic conditions, sufficient condition of introspection.

AMI defines introspection as a kind of immediate awareness of our own mental states which can provide us with justified, non-inferential judgments of these states. Hence, several philosophers construe introspection as a sort of acquaintance in which we are directly aware of our own mental states (Descartes 1637/1980; 1641/1986; Russell, 1912; Chalmers, 1996, 2003, 2010; Gertler, 2011, 2012). For a concrete view of this position, Gertler (2011) may serve as a representative for the acquaintance model.

Borrowing significantly from Russell's theory of "acquaintance" (1912)—i.e., that "we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths" (p. 73),—Gertler (2011) claims that “[b]eing acquainted with a mental state is by definition a direct (non-mediated) relation” (pp. 96-7). That is, introspection comes along without any mediation, and without its yielding a non-causal awareness of one’s mental life. We are “directly aware of” an ongoing mental state irrespective of our “awareness of something else”; for

---

51 Some versions of ISMI, also called simple accounts, claim that introspection proceeds simply and cognitively at a low level without involving additional mechanisms or systems. In contrast, pluralist accounts claim that introspection is dependent on various mechanisms or systems (see my Ch. V).
example, when “I learn that it's raining by hearing a radio weather report,” says Gertler, “my awareness of the rain depends on my awareness of the weather report and is therefore indirect” (p. 88).

Being "directly aware" also involves being present in the mind and not by virtue of any mediating object. So claiming that introspection is direct is on a par with claiming that it is immediately present to one's mind by virtue of having a mere mental state. For Gertler, introspective awareness is “metaphysically immediate” just by way of one’s having a mental state with a particular qualitative character—e.g., a sensation such as “the feel of pain” (p. 136). In other words, the claim that introspection is the direct awareness of having an experience involves one’s having an acquaintance with that particular experience.

Gertler's last point also echoes Russell's view: "It is… only what goes on in our own minds that can be thus known [or grasped] immediately" (p. 77). Gertler (2011) adds that "mental things are the only non-abstract objects we are acquainted with" (p. 91) and exemplifies that "having a certain kind of visual experience… involves… a kind of mental object… [a sense impression or a] 'sense datum'… [i.e., an immediate object of your awareness. But] in seeing [a non-mental entity, e.g.] a table before you, you are directly aware of sense data… [not of material objects such as] tables…" (p. 89).\footnote{No defense of a sense data view is at stake; the point only serves to illustrate a connotation of "direct."}

For Gertler, introspective awareness "is supported by a metaphysically direct, non-causal, relation of acquaintance to its objects." Such a relation is thereby immediate and possesses two primary features: a strong justification and an epistemic security (pp. 94, 87). But what do both a strong justification and an epistemic security amount to?

Let us begin with justification. When we are introspectively aware of having or being in a mental state—or of having the instantiation of a particular property of a state—we tend to generate a judgment to the effect that a state with a certain qualitative character is present. For instance, to judge that a “pain is present here or [that] pain is instantiated in me [implies self-attributing the sensation]: I am in pain” (p. 94). Gertler claims that "[i]nsofar as the metaphysical relation of acquaintance enables one to directly grasp… [the particular state and its] property, the resulting judgment will… be… justified" (p. 125).

That is, introspective judgments based on such a relation can achieve strong justification because of both the presence of the mental state which justifies the introspective judgment that I am in that state,
and *my awareness* of that particular state as one of being present—i.e., that I am having an experience or that that experience is occurring to me.⁵³ Importantly, according to AMI, I judge or form judgments on the basis of a certain justification of evidence without depending on causal relations.

An additional feature underpins the aforementioned justification: a “correspondence” between introspective judgments and occurring mental states. That is, when our awareness that our judgment of a “pain being present” corresponds to the presence of pain as an occurring state and not of something else or of anything beyond our mind, we are acquainted not only with the state at issue and the judgment on the presence of that state, but also with a “correspondence between these” two items (p. 100).

How does direct introspection capture such a correspondence? According to Gertler, such a correspondence is given rise by one’s “conceptualizing” the occurring state with its specific properties or with how an experience appears or feels to us. Also, the use of “demonstrative references” can assist us in justifying introspective judgments,—e.g., the ostensive “this” refers to the particular quality of an actual sensation, such as that of a specific pain. We can be aware of the presence of such a pain by judging “I’m now experiencing this quality” (p. 118; cf. Gertler 2001; Chalmers 2003). Since that demonstrative reference is supported by our being introspectively aware of the occurring pain, we then grasp the qualitative character at issue as something being *painful* and thus form the judgment that *this is present*.⁵⁴

For AMI’s advocates, the pain itself or the instantiation of the property of pain is considered part of the aforementioned introspective judgment. Namely, we are aware of it by demonstratively referring to that experience and by grasping its qualitative character. Thus we are aware of what it’s like to be in pain.

Additionally, Gertler endorses Chalmers’ view (2003) that in the mere grasp of “phenomenal qualities,” “direct phenomenal concepts” arise or “are incorporated” when an introspective individual apprehends the quality of an experience and forms a related concept (p. 235). This view similarly explains

---

⁵³ The claim that an individual is introspectively aware of a mental state in virtue of having the introspected state (Shoemaker, 1996) echoes a *primitivist view*: being in a state seems sufficient to put oneself in a position to know that one is in that state. For Gertler, it is not the mere presence of a pain what makes the target of introspective awareness a pain, but the judgment that accompanies the pain as that which is present.

⁵⁴ I bypass *expressivist* views, e.g., Wittgenstein’s denying that utterances like “it hurts” really attribute states to the individual who expresses it. For difficulties on expressive self-attributions, inferential connections, learning of expressions, symmetry between present and past tense uses, see Wright, et. al., 2000; Bar-On 2004.
the individual’s awareness of the correspondence between how mental states appear to her and how that awareness incorporates the judgment and its associated “phenomenal reality.”

As for epistemic security, the advocate of AMI promises that the direct relation of introspection with its target mental states achieves epistemic security if certain conditions obtain: if the introspective individual is able to offer the above-mentioned “justification” of her ongoing mental state provided that she pays “attention” to the occurring mental state with its qualitative character, if she is “scrupulously cautious” of it, and if she applies the corresponding “conceptualization” about it (p. 111). Some of these are what I have called the epistemic conditions of introspective access in Ch. I.

To illustrate, if you are aware of an occurring experience—a pleasant sensation you have while tasting a grasshopper’s taco—you carefully attend to it. If you are being scrupulously cautious regarding your awareness of that precise experience with its qualitative character—meaning that delicious spicy flavor and the crunchy texture you feel while biting it—and if you possess the cognitive skills needed to apply conceptual resources upon that particular experience according to how that experience appears to you and how you self-attribute the experience at issue (i.e., “pleasant,” “sensation,” “tasty” and so on), then your introspective judgment about such an experience is strongly justified and is likely to be epistemically secure.55 Nevertheless, if you were “to doubt” whether you are being introspectively aware enough of that experience with sufficient attention or care, then the justification of your judgment will “weaken” and “the degree of epistemic security” will decrease (p. 96).

4. Modeling a Compatible Account of Introspection

Previous sections have given us insight into the conditions of introspection. Those conditions determine the particular relation that introspection bears with its target mental states. Explaining the nature of introspection requires understanding that particular relation and understanding that such a

55 Whether applying conceptual resources to the character of states contributes to our recognizing instances of those states as being of similar character, and whether conceptual resources applied to initial experiences match and recognize other cases in subsequent experiences which resist interference of background beliefs, are issues that the acquaintance advocate does not explore. I shall turn to this point in my Ch. V.
relation is an important and controversial aspect of philosophical theories of introspection and requires examination.

The core of the problem posed by the nature of introspection can be seen by examining the central features of the leading opposing positions. There are those who claim that introspection is defined as an inner-sense mechanism. If introspection is defined as inner-sense, then it is causal—i.e., it is roughly how the relation that the introspective awareness bears with its target mental states is mediated. On the other hand, there are those who claim that introspection is defined as knowledge by acquaintance. If introspection is defined as an acquaintance, then it is direct—i.e., it is roughly how the relation that the introspective awareness bears with its target mental states is immediate.

With these prominent accounts broadly presented, we are now in a position to identify the problems for those views. Although these models have several problems, central difficulties concern the terms “inner-sense” and “direct.” For one thing, the definitions of introspection that both the inner-sense model of introspection (ISMI) and the acquaintance model of introspection (AMI) offer are too narrow. So although these models provide a number of reasons to adopt their views as the more plausible account of introspection, I argue that any suitable philosophical account of introspection must incorporate some aspects of both models. The consequence of this proposal would bring about a broadened notion of introspection.

I suggest that (a) the senses of the "inner" and the "direct" used by these models must be clarified. These clarifications will allow us to see that the models can be complementary and not mutually exclusive, as both causal and direct relations can serve to provide two different functional levels of introspection. Adopting a two-level view of introspection can show us a way out of the disputes between supporters of ISMI and those of AMI (§4.1 – §5).

(b) Rather than insisting that introspection strictly bears either a causal relation or a direct relation with its objects—in other words, that introspective awareness is exclusively either causal or direct—it should be agreed that the notion of introspection can incorporate refined aspects of both views. A conceptual distinction between a stimuli-induced introspection and a self-triggered introspection conceived of as modes of introspective access can be operative to bridge some of the gaps between the inner-sense and the acquaintance models (§5 – §6).
4.1. The Sense of "Inner-Sense"

I here clear up central difficulties that ISMI faces with regards to its term "inner-sense" and start drawing a definition suitable for a broader notion of introspection.

In debates on the mind the sense of the term inner-sense has been problematic. Folk-psychology typically equates "inner-sense" with "internal eye"/"internal observation," thus defining introspection. Given the inadequate analogy of a perceptual-like mechanism with introspection, this association has generated elementary confusions.

First, using "inner-sense" as applied to the introspective process is unpersuasive since no sensory organ has been assigned for introspection. Obviously, sensory modalities are operative via sensory organs over which we have a certain control, and those modalities function by individuating among sensory properties in relation with the environment. If introspection is construed as an inner-sense, one might think that it involves a sensory-based mechanism or a sensory system such as vision, audition, gustation, olfaction, somatic sensation, or even a vestibular system such as proprioception. However, it is implausible that introspection would involve any sensory mechanism.

In fact, neuroscientists and philosophers agree that no "sensory organ that takes brain events and/or mental states themselves as input has been identified" and “there is simply no empirical psychological basis to support the idea of a real tangible, inner perceptual faculty in human organism” that could be identified with introspection (Butler, 2013, p. 17; cf. Shoemaker, 1994). Also, the advocate of ISMI claims that no sensory organ is assigned for introspection—e.g., "when we are aware of happenings in our own minds, there is nothing that we are aware with" (Armstrong, 1993, p. 325).

If it is agreed that no sensory organ is required for introspection, but the claim that introspection is an inner-sense still lingers, it is reasonable to inquire whether introspection then possesses a derivable structural characteristic with a perceptual-like mechanism—and, if so, what it might be? If introspection were an inner-sense, it would be a process we would automatically undertake in the same way we apprehend objects of the environment. But introspection does not occur all the time or even regularly, as mere perception or thinking processes do.

Second, some might still wonder whether the alleged definition of introspection as an inner-sense or that which is analogous to perception may refer to their objects of apprehension. However, although
both introspection and perception are considered sources of knowledge, and in particular, sources of non-inferential knowledge, they do not require specific premises about ongoing mental states, modus ponens or rules of inference (cf. Clark, 2012), and there are important differences in terms of their objects of apprehension.\textsuperscript{56}

Recall that introspection arises in virtue of certain mental entities; whereas perception gives rise in virtue of non-mental entities—that is, the objects of introspection are psychological states and processes as opposed to being physical objects of our environment and its properties, which are the typical contents of perception. Introspection entails the mind’s awareness of itself or, in other words, introspection is object-involved; it involves both a mental state and a process.

The claim that an awareness of different entities entails differences of certain sorts seems uncontroversial—i.e., mental entities or properties seem to differ in fundamental ways from non-mental entities. Philosophers have emphasized the mental condition, such as “introspection is the mind’s apprehension of itself”; whereas perception is “the apprehension, by the mind, of [a] thing other than itself” (Mandik, 2010, p. 87). Moreover, psychological states and processes “do not persist through time as single, isolatable objects of perception... and other ordinary objects... [so, it is] difficult... to say when and where they begin and end, and to say how they are separable from and/or related to, one another on the basis of their observable characteristics” (Butler, 2013, p. 240). Most importantly, as mentioned above, the advocate of ISMI upholds the mental condition, too—that introspection is “information or misinformation about the current state of our mind” (§2).

This mere fact distinguishes it from perception. That is, in the perception of $p$ the relevant point lies in the relation between the mental state and a non-mental entity or an object of the environment ($p$); whereas in the introspection of a state of $p$—that is, my perception of $p$—the relevant point lies in the relation between that perceptual state and the introspective judgment that accompanies it.\textsuperscript{57}

Some philosophers seem to treat the aforementioned analogy by appealing to the same mechanism wherein introspective awareness and perceptual awareness are to share in the same

\textsuperscript{56} For reasons adduced in Ch. I, I don’t take introspection as knowledge or self-knowledge, but as a source of knowledge—specifically, as a cognitive process which can eventually lead one to achieve knowledge. Determining how different introspection is from other ways of knowing is another question that I do not examine here.

\textsuperscript{57} I bypass discussion on accounts of perception; see Dretske, 1999; Crane, 2005; Siegel, 2010; 2017.
cognitive resources to access their objects, or use the same resources that explain, one way or another, first-order experiences (cf. Prinz, 2004).

Although this claim is not uncontroversial, it does not follow that because both perception and introspection entail certain types of awareness, these processes use the same cognitive resources for apprehending their corresponding objects such as mental entities as opposed to non-mental ones. Neither does it follow that these processes work in a similar way (cf. Shoemaker, 1996, III; 2012; Armstrong, 1993, Ch. 15, §II) irrespective of differences in terms of their objects of apprehension. Nor does it follow that introspection is usefully identified as a form of perception or quasi-perception—as I will show in the next section. Brain distinctions have been found when individuals undertake introspective awareness vs. perceptual awareness (cf. Fleming et al., 2010).

Third, in the absence of a sensory organ and a similarity between objects of apprehension within which we could identify introspection with a perceptual-like mechanism, it seems natural to reject that there is “something that it is like” to be in introspection or that the aforementioned analogy lies in a distinctive phenomenology. Moreover, it is reasonable to claim that introspection construed as "inner-sense" does not entail a distinctive phenomenology, nor is there “something that it is like” to be in introspection. It is unlikely that the sense of "awareness" involved in introspection is non-cognitive or that it entails phenomenal properties.

Even if introspection is considered a kind of “monitoring consciousness” different from both “phenomenal consciousness”—i.e., the subjective experience or what it is like for one to undergo certain mental states—and “access consciousness”—i.e., the ability of information processing available for discrimination, reasoning, speech, and action control (Block, 1997; 2007a; 2007b)—the mere use of “consciousness” as applied to introspection is problematic. Notice, for example, that it is unlikely that the problem of introspection entails the well-known “hard problem” of consciousness (Chalmers, 1996).

Minimally, while there is “something that it is like” to undergo a perceptual experience of a certain object with particular properties, no phenomenology or appearance of objects with phenomenal character seems to be in play in introspection or to be part of introspection. The typical example of the red tomato used in the literature may illustrate the point. While there is something it is like to see an object such as a tomato, since a distinctive phenomenology of redness and roundness is at stake—i.e., the properties that
my visual experience represent the object as having—no distinctive phenomenology is involved in introspectively judging with regard to my visual state that tomatoes are red and round. Being introspectively aware of a mental state does not seem to entail experiencing qualities or a "what it's like."

Besides, philosophers argue that no distinctive phenomenology is involved in introspection. We are not introspectively aware of "beliefs and thoughts by having sensations or quasi-sense-experiences of them" (Shoemaker, 1996, p. 207). Even though imagery may be at issue while introspecting—e.g., while some "thoughts are associated with sensory images of what they represent," others are accompanied by "subvocal speech"—"there is no distinctive feel of introspection" (Prinz, 2004, pp. 51-2). Further, "[w]hile some sort of phenomenology may accompany the introspective process, any phenomenological feel associated with introspection [itself] is strictly incidental and does not help to epistemically ground an introspective [judgment]" (Gertler, 2011, pp. 134-5).

Despite consideration on certain mental states (commonly identified as conscious experiences, entailing phenomenal character or an intrinsic qualitative property such as a pleasant sensation while tasting a taco), no phenomenal constituent is typical of introspection. Simply, when by introspection you judge that tasting a taco is a pleasant sensation, the phenomenal character that that experience bears belongs to the sensation itself, not to introspection or the introspective judgment. So it is relevant to distinguish introspective target mental states having certain phenomenology—either type-states such as gustatory sensations or token-states such as tasty or spicy sensations—from the introspective process per se that bears a certain phenomenology.

The obvious conclusion, then, is that introspection can be considered merely intentional, and it uses mental state as targets. So to claim that some target states of introspection entail a phenomenal character does not mean that a phenomenal or qualitative character is a constituent of introspection. While such a character involves the state available to be introspected (which could be a mere conscious experience with a particular content), it is by introspection that a judgment of that very experience can apply to both the "tasty" and the "delicious" predicates for the character of the sensation and the self-attribution of the experience at issue.58

---

58 The structure of conscious experiences in accordance with distinctive phenomenology—i.e., qualitative character and its corresponding information—can be determined and organized by introspection, or so I
Therefore, introspection does not exhibit the hallmarks of a perception-like mechanism, strictly speaking. Or, introspection does not at least possess a sensory organ, since the type of awareness of mental entities differs from apprehending non-mental entities, and no distinctive phenomenology is likely to be involved.

Importantly, ISMI itself rejects the claim that introspection entails a distinctive phenomenology as well. If introspection were to involve a distinctive phenomenology, it would not be definable in physical terms—i.e., it would be “physically irreducible” or would not be “accommodated in the physical domain” (Goldman, 1993, p. 24), which goes against the goal of ISMI.

To regard introspection as an inner-sense in virtue of the aforementioned features is untenable. Yet ISMI’s goal is to show that introspection can be definable in causal terms or according to the way it works, based on causal-relational properties.59

4.1.1. Some Consequences of Advocating for an Inner-Sense View

I here show that by construing introspection in causal terms, ISMI faces at least one central problem which constrains the range of its merits. This problem concerns a reductionist view of introspection. The consequence of this view undermines the peculiarity of introspection.60 It is not my intention to cover the implications in depth, much less to pursue the controversies concerning the problem.

Before I proceed to discuss the problem, a question arises: is it possible to preserve the causal role of introspection without taking “inner-sense” at face value? Since we have already cleared up the conceptual confusions, the answer must be positive. Armstrong himself contends that introspection is a

59 For additional objections to the inner-sense view, see Shoemaker, 1988; 1994; Nichols & Stich, 2003.
60 Although “peculiarity” is sometimes (erroneously) construed as a sort of “privilege” in the literature—i.e., enjoying either a first-person authority about our mental lives or a range of attributions provided by a first-person’s method—I distinguish “peculiarity” from “privilege.” Privilege access refers to our “epistemic position vis-à-vis propositions ascribing current mental states [or processes] to [ourselves]... [This access] is favorable [or authoritative] in a way no one else’s position is” (Alston, 1971, p. 230). Thus, resulting in an asymmetry of first-person and third-person—i.e., we do not know our minds in the same way in which we know about the minds of others (cf. my Chapters I, II, III). Peculiarity is also identified as the difference thesis defended by some philosophers in the contemporary literature in which “introspection is different from other ways of knowing about the world” (Smithies & Stoljar, 2012, p. 4).
perceptual-like mechanism because perception and introspection are defined in causal terms or are determined by a causal connection with their respective objects of apprehension.\textsuperscript{61}

Arguably, then, the problem to solve is to determine whether introspection and perception are structurally similar simply because of the causal way in which these mechanisms work. While perception and introspection are different sources of knowledge, if we appeal to the causal role, they seem to work in parallel.\textsuperscript{62}

To elaborate, perception is ordinarily said to be a causal process in that to perceive \( p \) is to be causally affected by \( p \). As you are about to eat a taco, you see a grasshopper on it (let’s leave alone whether the grasshopper is already fried or is still jumping on the taco). There is a visual perceptual state \((s)\) with a particular content \((p)\) and a distinctive phenomenology \((\phi)\)—that is, you are seeing something red having the shape of a bug with enlarged hind legs. Assuming that \( s \) is veridical, it accurately represents \( p \)—i.e., that particular bug on your meal. Since there is a grasshopper on your taco, it contributes to the causing of your having such a visual perceptual state. In other words, your perceptual state of that object depends on a certain causal relation, or your state is causally connected to that specific object.\textsuperscript{63}

Similarly, the advocate of ISMI contends that introspection of a mental state is to be causally affected by a particular state. Introspective judgments in particular are caused by occurring mental states yielding self-attributions of those states. If we take into account both that introspective judgments are ordinarily defined as statements about one’s own current mental life, and that in self-attributing one judges or makes a judgment about one’s mental life, it would be natural to presume that, following the last example, introspective awareness of a visual state of a grasshopper on your taco would bring about a judgment on which that particular state with its distinctive phenomenology is based. Specifically, there

\textsuperscript{61} Despite Lycan’s wrongly characterizing introspection as a quasi-perceptual process (1996), he correctly distinguishes between an introspective target state and an introspective judgment, and the notion that two entities can be causally connected, but not constitutively connected.

\textsuperscript{62} How different their causal relations are is another question that I do not examine here.

\textsuperscript{63} I bypass discussion on whether (a) we perceive physical objects via a direct relation between the perceiver and the perceived object, or perceive physical objects indirectly via mental representations of physical objects; (b) our perception is more reliable than introspection, or both processes are reliable at the same level; (c) the “contents” or the specific kind of representation that perceptions and introspections entail; (d) how representations are stored and what properties these bear.

110
would be a self-attribution that such a visual experience is happening to you and/or the generation of an introspective judgment that *I am seeing a grasshopper in my taco.*

In detail, introspective awareness would be caused by the mental state it represents with its distinctive *ph* and respective to *p*, considering that the content of your visual state is actually a grasshopper on a taco instead of a maguey worm in a drink. According to ISMI, additional outputs may also play a role. The phenomenal character of your introspected visual experience can figure in the causal network by manifesting external behavior such as bodily sensations—e.g., disgust at eating insects or a gag reflex. It may be the case that those responses can also induce further introspection, generate different judgments that accompany the initial experience, or enroll in subsequent experiences. It may be the case that a cluster of states such as fear or a belief about being poisoned by eating live or dead bugs can arise as well.64

This way, then, the relation that introspection bears with its target mental state is causal. The advocate of ISMI claims that *causal* is a necessary and sufficient condition of introspection when the three basic conditions obtain. So if mental states necessarily *cause* introspective awareness, and if introspection is causally mediated in a similar vein in which physical objects cause perception, it follows that introspection is a *perceptual*-like mechanism. However, we cannot take this at face value, since not only are serious differences in play, but a crucial problem most importantly occurs.

Roughly, as AMI’s advocate argues, “[t]he difference concerns how mental states are represented. Just as no one other than you can refer to you with the word *I*, no one else can represent your mental states in precisely the way your scanner represents them.” Accurately grasping a non-mental entity or an actual object involves (perceptual) awareness of it only indirectly via a mental representation, whereas accurately accessing a mental entity involves (introspective) awareness of the mental representation itself (Gertler, 2011, p. 145, my emphasis).65

As for the crucial problem of ISMI, notwithstanding the above-mentioned points, the core of ISMI remains unpersuasive for the advocate of AMI. That is, defining introspection in causal terms is

---

64 Insofar as I know, this point has not been explored by ISMI. My model offers this possibility (see Ch. V).
65 Different versions of this model may concern whether we can be directly aware of *all* our mental states or *only* some (Gertler, 2011, p. 95). For objections concerning assigning phenomenal character to mental states, see Schwitzgebel 2008, p. 263; Smithies & Stoljar, 2012, pp. 10-11; Ch. 4, 15; Gertler, 2012.
problematic for the rival position, since the description entails physical features or neural connections even if such connections are not accessible through introspection (see Gertler, 2011, Ch. 5).66

By offering a causal-role model of introspection, mental states correlate to neural states in terms of their operation, their connections between systems according to their causal networks of inputs and outputs, or, more specifically, their exertions of influence between their sensory inputs and behavioral outputs. If ISMI could define introspection in those terms, and if it incorporates the mental to the physical, introspection would be considered a physical process, and, therefore, ISMI would face the problem of reductionism. Reductionism broadly construed is the thesis that psychological theories, properties, or entities identify with physical/neural theories, properties, or entities. A type-reductionism or a version of reductionism specific to this discussion is that introspection reduces to perception or a perceptual-like mechanism and/or its properties.67

To clarify, if introspective awareness is reduced to a perceptual-like mechanism, or if it is on a par with perceptual awareness, no epistemic distinctiveness between these types of awareness holds. That is, the peculiarity of introspection would be undermined. Recall that peculiarity is the epistemic thesis that introspection is a distinctive type of awareness or way of knowing about our minds—that is, of our mind’s awareness of itself. And there is a sharp contrast between the way in which we know or are aware of our minds from other types of awareness and ways of knowing about the world. You and I are not in a position both to be aware of our mental states and to be aware of the objects of our environment in the same way and/or under the same conditions.

It is worth mentioning that peculiarity does not imply superiority—i.e., the claim that introspection bears an epistemic status superior to that of perception or other cognitive mechanisms. Additionally,

66 Whether this view entails a type-physicalism concerning qualitative properties as identical with neural properties, and whether the specific physiological mechanism that this particular model addresses has already been identified in cortical areas yet, each goes beyond the present purview; cf. Goldman, 1993.
67 Contemporary philosophical theories have provided a number of responses against various reductionist programs (e.g., Kripke, 1980; Jackson, 1982; Levine, 1983; Chalmers, 1996). Examining these responses is far beyond the scope of this project. I bypass other forms of reductionism that may play a role here.
claiming that introspection is peculiar is not to suppose that it is the unique kind of awareness or way of learning about our minds.\textsuperscript{68}

The consequence of reducing introspection to a perceptual-like mechanism is that no peculiarity of introspection would be preserved. So far, the problem is serious; introspection should not be reduced to a perceptual-like mechanism since one of the prerogatives of introspection is its peculiarity—that is, introspection’s distinctive type of awareness. The question (now) is whether or not there is a way in which we might still preserve the causal condition without facing all the problems of the ISMI. We will see that this is possible at §5.

4.2. The Sense of "Direct"

In clearing up central difficulties that AMI faces with regards to its term \textit{direct}, I here draw a definition suitable for a broader notion of introspection.

Acquaintance theories are generally construed as theories of knowledge or self-knowledge, and roughly distinguish “knowledge by acquaintance” private and subjective \textit{from} “knowledge by description” public and objective. Although these views lend themselves to some light-shedding on the nature of introspection, a number of difficulties arise with the epistemic sense of “direct.” I discuss only two of those difficulties. One concerns the reference to consciousness and the other concerns conceptualization and background beliefs.

Before proceeding to the discussion, it is important to be clear that although introspective judgments about mental states can be considered \textit{true} according to some philosophical views, what is at stake in the current discussion is neither the truth value of introspective judgments nor the endorsement of \textit{ideal} or perfect forms of privilege access to our mental lives. For what had been considered “true” had come only as a result “of a reliable process, causally linked to its truth maker—namely the presence of the self-attributed [mental state]” (Smithies and Stoljar, 2012, pp. 12-3). That is not relevant here. In a nutshell, direct awareness of mental states does not involve infallibility of introspection (cf. my Ch. I-III).

\textsuperscript{68} I have defended an asymmetry between first and third-person perspectives (cf. Ch. I-III). Armstrong (1993), however, claims that such asymmetry is contingent, since our inner-sense can be linked to the state of another individual via telepathy (p. 124).
First, in the philosophical literature the term *direct* sometimes refers to "subjective character," "consciousness," or "feeling," as opposed to *non-conscious* or *indirect*. At some point, the AMI advocate will claim that introspection involves a certain phenomenology. "In acquaintance, direct awareness of a mental state involves a particular sort of phenomenology"—e.g., by being introspectively aware of a property of certain sensation, say, "pinching [you are acquainted] with that property directly [which] involves its phenomenology, namely, that slightly uncomfortable feeling of pressure. And this phenomenological aspect of introspection is crucial to [your]... grasp of the phenomenal property" (Gertler, 2011, p. 134).

Arguments against a distinctive phenomenology of introspection have been offered (§4.1). This issue certainly puts a constraint on the range of AMI merits.

Second, for AMI, when you are introspectively aware of a perceptual experience, the experience does not causally engender your introspection. Nothing mediates between your introspective awareness of the mental state and the mental state itself. That is, there is no metaphysical gap between that state that is being grasped through introspection and the state itself—i.e., your particular painful sensation.

Moreover, the expression "nothing mediates" means that introspection does not have to do with an observation of behavior which is in any sense other than that from which we can infer that we are in a certain state. You do not need to infer the presence of a state from evidence. In introspecting the sensation you are having right now, you seem able to grasp its basic nature directly—that is, through its particular quality—whereas you otherwise seem unable to achieve this same sort of grasp of anything that is physical.

Even if we grant that the non-causal relation obtains, AMI still needs to explain whether the role conceptualization has while we are introspectively judging a particular experience would not be considered a sort of mediation. That is, it must show how introspective awareness, in spite of its being direct and immediate, still grasps the state as a state of a certain kind by means of conceptualization. In other words, "direct" introspective awareness would be *mediated* by conceptualization.

It is likely that conceptualization imposes certain mediation between the state and the introspective judgment. Here, conceptualization would work as a mediation of mental states, given that it is not part of the mental states themselves.
Although Gertler (2011) acknowledges this difficulty, she argues that conceptualization works by our distinguishing and classifying a mental state as of a particular kind “by its epistemic appearance”—e.g., you can distinguish a pinching sensation from a tickling sensation according “to how your experience feels to you—the phenomenal quality it epistemically appears to exhibit... to introspection.” That is, you conceptualize a property of your sensation or the only discernible aspects of its phenomenology “by using your grasp” of how the experience seems to you or “how it epistemically appears” based on the way you are applying conceptual resources in the judging of what such a sensation with a particular character is doing to you (pp. 114-5).

Now, how direct introspection of an occurring experience works depends on some epistemic conditions of introspective access (cf. §3, and Ch. I). That is, it depends on whether the introspective individual is paying attention to the experience in question and is exercising caution in grasping its particular feeling. In turn, that “attention”/“exercising” move depends on the way the state is appearing to her. In being directly aware of that state, the advocate of AMI will claim that “your sensation epistemically seems to involve a [particular] quality” and “no causal process mediating between the [phenomenal] reality and its epistemic appearance” is in play (pp. 95; 112-5; 119). When you are introspectively aware of a qualitative state, “there is no appearance/reality gap” (Hill, 1991, p. 127); so conceptualization is not mediation.

As for background-beliefs, one also might inquire whether these would not mediate the introspective judgment and self-attributions of mental states, since introspective judgments may contain such background-beliefs that causally contribute to the formation of judgment. For example, you can reach an introspective judgment about making your mouth water by seeing a grasshopper on your taco. Clearly, part of that judgment is about the distinctive phenomenology corresponding to the occurring state.

Some philosophers claim that formation of introspective judgments can be influenced by background-beliefs as well as by proprioception, expectations (Schwitzgebel, 2012), or motivations (Nisbett and Wilson, 1977). However, Gertler asserts that even if background-beliefs can causally contribute to the formation of judgments, these do not interfere with the justification of introspective judgments which is given by a direct awareness of occurring mental states. That is “exclusively
determined by how things seem phenomenally ...Success in this endeavor will neutralize the influence of background beliefs” (Gertler, 2012, p. 110).

Even if we grant that those beliefs can influence judgments, cognitive scientists argue that these are not always corrupting sources or distractors from the ongoing introspected target state, since it is possible “to eliminate noise” (Reyes and Sackur, 2014; personal communication).69

5. Two Functional Levels of Introspection

We have seen some of the problems that the inner-sense model of introspection (ISMI) and the acquaintance model of introspection (AMI) face. As mentioned, their definitions of introspection are too narrow; they exclude important aspects of the rival position, or their respective assessments too simply imply that the rival approach is irrelevant to an understanding of the nature of introspection. Armstrong and Gertler commit themselves to a false dichotomy between the ISMI and the AMI since, for them, only one of the aforementioned conditions or types of accesses, either causal or direct, is conducive to introspection or is defined as introspection.70

Those philosophers have been so embedded in discussing whether introspection bears a causal or a direct relation with its objects that they overlook the possibility that different functional levels might be at stake. Both causal and direct can be justified as two different functional levels. So rather than insisting that introspective awareness is either causal or direct, it should be agreed that the two levels of introspection hinge upon their causal roles and their non-causal roles. This alternative view naturally invites a different assessment.

In searching for a definition of introspection compatible with both positions, I argue for the possibility that the ISMI and AMI complement, not exclude, each other. Both approaches to introspection can coexist and assist each other. Introspection can be explained in physical terms and can also be explained in metaphysical terms, with some limits of the former being fulfilled by the latter and vice versa. Thus, this suggestion involves the possibility of accounting for introspection in accordance with its causal

69 In Ch. V, I consider background information as a variable of introspection.
70 For objections to inner-sense theories, see Shoemaker, 1996; Gertler, 2011; Butler, 2013. For objections to acquaintance theories, see Horgan & Kriegel, 2007; Stalnaker, 2008.
role and its non-causal role, depending on the aspect or property to be examined. Indeed, there is no way of proving that the accounts of ISMI and AMI are as independent as the debate assumes.

Let us see how this suggestion might work by presenting three ways of thinking about it. First, even if we grant that introspection bears a causal relation with its target mental states, such a causal relation is fundamentally distinct from perception, or does not work as it does in perception. Although ISMI retains a certain appeal in that introspective awareness—or an introspective judgment and its target mental state—can be causally connected, it is likely that such a causal condition is contingent. Alternatively, such a causal connection might not be the unique way to engender introspective judgments or to initiate introspection; it seems to leave out other explanations about introspection connected to its privileged access.

But the other way around seems to apply, too, even if AMI seems a promising model of introspection, since "direct" cannot capture the condition that introspection requires or all that there is to introspect. It seems to be insufficient in providing a definition of introspection or in understanding the nature of introspection. So causal and direct comprise either contingent conditions or fall short as independent definitions of introspection.

Second, if introspection is causal, then the direct connection is neither a necessary nor sufficient condition for introspection. If the direct connection were only to minimally contribute to provide a metaphysical explanation of introspection, then we can be almost certain that this view would be useful in shedding light upon the nature of introspection or some aspects of it. Similarly, if introspection is direct, then the causal connection is neither a necessary nor sufficient condition for introspection. If the causal connection were to only minimally contribute to providing a physical explanation of introspection, then we can be almost certain that this view would also be useful in shedding light upon the nature of introspection or some aspects of it.

Three, the causal condition is by no means as negligible in introspection as the advocate of the acquaintance view supposes. On the contrary, it significantly explains our notion of introspective awareness at a certain physical level or from a naturalist approach. Nonetheless, introspection cannot be reduced to this level, since there is yet another level of introspective awareness that is important to preserve. In other words, we can explain how introspection works in physical terms, such as being part of
a causal network; or in how it might entail a sense of causality in which its elements exert influence; or in how it can be defined as the result of a causal relation with its target mental states. Additionally, without involving any conflict, we can explain how introspection works in metaphysical terms as an immediate awareness that capture those aspects which resist physical reduction, or which are difficult to explain in physical terms, such as the first-person perspective, the so-called “presence” of a state, or the putative way the mental state epistemically appears to the individual and its connection to phenomenal reality. These explanations simply do not compete with each other.

Alternatively, we can explain how introspection works according to neural properties already identified by the recent use of brain-imaging techniques, and how some of those properties can be described using non-phenomenal facts. Additionally, without involving any conflict, we can account for the qualitative character of the properties of mental states. Again, these explanations do not compete with each other.

Thus presenting a non-exclusive alternative view which can accommodate both causal and direct relations may be more attractive than the previously established view, despite not being the upshot that ISMI and AMI attempt to provide. In other words, in order to account for the nature of introspection, the two different functional levels that involve the causal relation and the direct relation must be considered. Neither of these views provides conclusive evidence, even though introspection can be explained at a physical level, at least according to recent scientific findings on the neural bases of introspection (Fleming, et. al., 2010).

This proposal, then, will be modeled on the framework of two different functional levels of introspection according to how this cognitive process shifts direction upon modes of introspective awareness (see §6). Thus, I offer the possibility of enhancing the notion of introspection by considering a proposal which, if adopted, may lead to reconciliation between those who claim that introspection is defined in terms of a causal relation and those who claim that it is defined in terms of a direct relation.

As far as I know, neither ISMI nor AMI—rather, neither Armstrong nor Gertler as representative thinkers of each camp—manifestly recognizes that a combined notion of introspection might be plausible. The upshot of this proposal would bring about a qualified notion of introspection.
6. Modes of Introspective Awareness

I have suggested that ISMI and AMI are not mutually exclusive but indeed are actually complementary. In this section, I show that both viewpoints can be reconciled while broadening the notion of introspection. In particular, I propose that inside a qualified notion of introspection there is room to distinguish between (what I call) stimuli-induced introspection and (what I call) self-triggered introspection. I argue that this conceptual distinction may eliminate a false dichotomy between ISMI and AMI which claims that only one condition, either causal or direct, is conducive to introspection or is defined as introspection.

If the false dichotomy between ISMI and AMI is removed, a qualified notion that accords with (what I call) different modes of introspective awareness can be appreciated. The labels stimuli-induced introspection and self-triggered introspection have been chosen to illustrate how the rival positions could be bridged if we consider modes of introspective awareness instead of only one putative definition of introspection or the definition of introspection.

Stimuli-induced introspection is, in a certain way, the instantiation of the causal property of introspection. Or at least it may plausibly fill the role of the causal relation with its target mental states. And self-triggered introspection is, in a certain way, the instantiation of the direct property of introspection. Or at least it may plausibly fill the role of the direct relation with its target mental states. We can further get a grip on the distinction between these models of introspective awareness.

Since introspection has been given as a function of its target mental states, the differentiation between modes of introspective awareness also can be presented as a function of their objects of apprehension. The modes of introspective awareness hold ceteris paribus assumptions; we can presume that basic conditions of introspection and epistemic conditions of introspective access obtain. But no further aspect of a similarity between ISMI and AMI that has not been examined in this chapter is intended.

The taxonomy that I offer can usefully capture modes of access that shed light on the nature of introspection. However, in the introspective process or current event itself these modes of introspective awareness are close to being indistinguishable. In other words, how an introspective individual is able to identify or to discern by way of introspection whether her mode of awareness is stimuli-induced or self-
triggered is a question that I do not examine here. The modes of introspective awareness that I offer are intended to serve as a theoretical alternative to understand the nature of introspection.71

In its simplest form, I take introspection as a way to be aware of our own current and recently past mental states, and to self-attribute those states. The sense of awareness at stake alludes to a faculty of self-probing our mental states together with our abilities to form judgments of those states, and/or to report them accordingly—i.e., whether a specific mental state with certain qualitative character can be manifested thus and so, or whether it merely seems thus and so from our first-person perspective (cf. Ch. I).

In terms of its scope, I take introspection as the span of information-processing that extends from having a mental state, such as being hungry—i.e., the target of introspection—through to its passing from the self-attribution of the state “I am feeling hungry” and then on to its deriving from it a corresponding judgment—that I am in a state of hunger—which can then be verbally reported or not. Depending on the introspective capacity of the individual and on other conditions, whether reports can come in levels and degrees instead of yes or no responses, and whether the introspective individual executes self-ascriptions exactly when the experiences occur or not, go beyond the present concern.

6.1. Stimuli-Induced Introspection

According to this mode of introspective awareness, a mental state induces the process of introspection and brings about an introspective judgment. For simplicity’s sake, I capture only the judgment as induced by a mental state—that is, I take for convenience’s sake only a fragment of the introspective process to explain the point. Let us consider the first mode of introspective awareness in its simplest form:

Stimuli-induced introspection (SII): a passive or receptive process whereby some specific mental states (mental stimuli) can spontaneously or automatically cause introspective awareness.

Consider this mode of introspective awareness as some kind of involuntary process. Notice that the stimulus is not caused or induced by an external source of the physical world, as happens with

71 It can make a difference in terms of the understanding the operation of introspection, its development, and its interactions among mental phenomena—or variables. See Ch. V.
perception, but is a mental-state-induced introspection. However, an introspective judgment can be caused by a mental stimulus connected to an object of the environment—i.e., a sensation result of a relation to a non-mental entity—or by a combination of mental stimuli. Set aside that sometimes a stimulus is so strong or threatening that it prevents us from introspecting the target state—e.g., an intolerable pain.

A bodily sensation and a feeling of displeasure while receiving an apathetic and loose hug induces your introspective awareness of those target states, and you form the judgment: “I hate this cold hug sensation; it is worthless.” (Call it introspective awareness induced by an American-style hug’s sensation).

Introspection here is induced by a mental state—which can be simple or combined—bringing about a corresponding judgment.

A yearning of permanence while you are getting along with the usual people and following your routine induces your introspective awareness of that state to the effect that you bring about a (contrary) judgment: “Really? Do I feel this? Well, I don’t desire that at all, actually I don’t fear my mutability.” (Call it introspective awareness induced by a desire of an everlasting quality).

Introspection here is induced by a mental state bringing about a contrary-effect judgment to the initial mental stimulus. Introspection does not necessary respond in accordance with the input; the output might be opposed.

A feeling of perplexity while you are opening an umbrella and I meanwhile collapse on my side—instead of my using your umbrella to cover both me from the rain—induces your introspective awareness of that state to the effect that you bring about the judgment: “I think I better get out of here!” (Call it introspective awareness induced by the “fainting goat” feeling).

Introspection here is induced by a mental state bringing about an unexpected output: a new state and/or an atypical judgment.72

72 The core of “fainting goats” is captured here: https://www.youtube.com/watch?v=we9_CdNPuJg. By “unexpected,” I mean outputs such a judgments or additional states that can emerge. Against this are
Additional refinements help to clarify how SII works. Notice that the character of the mental stimulus induces different introspective events and exhibits distinct types of judgments. Additionally, the introspective individual would be able to find out both what is bringing about the connection between a mental state and its introspective judgment and the additional mental stimuli serving as causes of introspection, beyond the ordinary representations of mental states.

If we grant, with the inner-sense view, that introspection does not have to tell us about the causal aspect of the experience or does not necessarily tell us about the causal source of your state (although it does not exclude its possibility of getting some of them by further introspection), but rather instead gives us enough information about mental phenomenon, and if we grant, with the acquaintance view, that introspection tells us how a particular experience appears to us, it would be reasonable to maintain SII as a plausible mode of introspective awareness. The relevant point here is that you are ensured that your being able to react from a/or induced by a mental state in particular then allows you to elaborate a judgment on the target mental state.

Now, we may wonder whether questions can also be considered stimuli of introspection. Although questions seem to be stimuli of a different nature, one might ask if being queried by someone to introspect when a stimulus attention appears qualifies as SII, since it is the goal of following instructions, not the stimulus itself that drives the introspection.

Consider a scenario in experimental psychology in which you have been asked by an experimenter to follow her clear instructions in order that you be able to report precisely what you are feeling when a stimulus \( s \) is being presented to you.

The experimenter knows—before you do—what \( s \) is, and what your possible experiences and reactions might be. She will evaluate your reports depending on facts, circumstances, and the specific situation, \( s \), the reception of \( s \), and the additional information she possess—and she will use your reports as useful scientific evidence.

When a specific task motivates an introspective event, or when the task entails direction or a guide to execute something or to follow instructions to perform something, or when the context and/or the

---

typical outputs, such as: when you hit me, I respond with pain, I shout or express anger; I hit you back or run away.
situation is specific to the extent that the experimenter is waiting for a response from you, the case can be plausibly considered SII.

Additionally, even if the event bears a certain pressure to perform the task within a frame of time and to respond the question at issue in a specific moment, you are aware of the expectations of the third-person—that is, the experimenter is waiting for your response, and she is following a methodology to deal with the information in the form of introspective reports about your experience. Therefore the case can be plausibly considered SII.

The question that arises here is whether there is a difference in the mode of introspective awareness when no third-person is guiding or supervising the progress of your introspection, and when no expectations and pressures are in play. Be that as it may, however, let us recall that the point at stake is that when there is a certain stimulus that induces introspection, a specific state's being particularly intense or a protruding process which pops-up is different from when there is effort and intention to undertake an introspective event.

Finally, even being introspectively aware of certain ongoing mental states, in the absence of any causal connection between those states and introspection, means that there is room for a non-causally mediated introspection; that is, since there is not always the case that introspection is induced by a stimulus, it is convenient to examine self-triggered introspection as another plausible mode of introspective awareness.

6.2. Self-Triggered Introspection

As always, you are rushing. You have to leave right away, but you prefer to stop for a minute to inquire as to what your own current mental state is. Not only do you choose to start an introspective event or to initiate a self-probing of your mental states, this mode of introspective awareness will also select the target mental state.

Contrary to stimuli-induced introspection (SII), self-triggered introspection (STI) actively selects a mental state from a cluster of occurring states or a variety of mental phenomena happening to you, or as they appear in your stream of consciousness.
Consider the previous examples as happening to you right now: you are being selective in the introspection of your current American-style hug’s sensation, in your affective psychological response to it, or in your desire for an everlasting quality or a fainting goat feeling. Alternatively, you can choose to probe the state that such a desire may hide or displace—e.g., a feeling of panic for an unavoidable situation. Let us consider the second mode of introspective awareness in its simplest form:

**Self-triggered introspection:** an active or selective process whereby the individual’s own interests or volitions initiate introspective awareness.

An example may help to illustrate how this mode of introspective awareness works. While being in the elevator, you initiate introspective awareness of a mental state. No doubt several states are occurring in your mind—some are passing, others have just recently passed, and so on; and numerous other examples could be easily mentioned. But let us focus on how specifically STI works.

You are in a good mood today for whatever reason; you feel motivated to smile and to make eye contact to share that positive state. But you begin to worry when everybody in the elevator seem to be *automata*; they are avoiding eye contact with you at all costs. You think they might be thinking that cheerfulness is contagious (and so you smile to yourself). But in trying to be comprehensive, you also believe that these people are so busy and overwhelmed with hundreds of their “to do lists.” Or, with a certain dose of concern, maybe you instead judge that they are in panic out of a fear of being accused of sexual harassment for having undertaken visual contact.

You self-probe a current mental state and inquire on its character: for example, what the specific qualitative character of feeling good is, and what the contents of such a feeling would be (call it the elevator’s feeling).

Although certain mental phenomena or aspects of your phenomenology are set aside because of unawareness, defense mechanisms, or repression—e.g., a desire of yelling to confront the automata—some of them might be accessed by further introspection.

Contrary to stimuli-induced introspection (SII), self-triggered introspection (STI) is a mode of awareness guided by will and effort, and it can be either inhibited or encouraged at will. This mode of introspective access is not caused by a stimulus—that is, no inputs or mental stimulus engenders the
initial introspective event. Instead, out of either curiosity or concern, it is activated or triggered by the introspective individual’s effort, genuine interest or intention to self-probe a current state with the implication of looking into her stream of consciousness. In a nutshell, the individual initiates introspective awareness and seeks her own mental life.

Additional refinements help to clarify how STI works. STI comprises two main interacting functions: a voluntary/intentional control one and a selection one. This mode of introspective awareness is accompanied by a volition and an intention to undertake a search or by simple inquisitiveness as one self-probes one’s mental life for the sake of investigating it. Also, STI is construed as an exploratory action meant to obtain information about one’s mental life, which can be in the form of a mere desire to look over and/or to learn about one’s own mental features. By self-probing a certain mental state here and now the introspective individual sets out to find out its character and contents and to distinguish among states. But exactly how the introspective individual is able to account for a state among several mental states is another point that I do not explore here.

The point at stake is that the introspective individual not only self-probes her mental life voluntarily, but she also points out and selects the mental state to be introspected. To be clear, I do not refer to her having any voluntary control over the state or her capacity to inhibit the process at issue, but I do refer to her being able to introspect what is there to be detected via experiencing. So in this sense, there is an active role being undertaken by the individual in this particular mode of introspective awareness. That is, STI is actively selecting among the mental states and depending on what aspects of her phenomenology the individual attempts to find out. For example, you may be more interested in mental states that manifest a richer stock of qualities or in those states that interact with others (see Ch. V).

In short, voluntary/intentional control ends up deciding between the states and the execution of introspection, and selection ends up determining the classification of the target mental state with its particular character to be introspected. Whether these functions start in getting direct control of the will, planning in the implementation of steps according to current contexts, moving on in gathering certain
information according to theoretical distinctions, or devoting important load/charge energy, among other things, all are part of a different discussion that cannot be addressed here.\textsuperscript{73}

One might add that STI appears to be dependent on volition/intention and so require a causal condition, too, since the individual’s own interest can also cause introspection. It seems this mode of introspective awareness is cognitively triggered by the volition/intention to access states, and it would be wrong to presume the absence of a causal network between states and introspection.

Additionally, one might add that introspection appears to be caused by decisions, deliberations, and the like. For example, when you introspect you have to make a decision about how to attain the best possible outcome under bounded conditions—meaning within the constraints being imposed by the availability of information, one’s cognitive skills, time, and so on. Hence you would necessarily be a “passive” introspective individual of your mental activities. Also, the mere fact of an active and not a passive process implies the functioning, performing, or causing of the action. So emphasizing what initiates the process creates pressure to address the problem of causality again.

Although it is not easy to reach a consensus as to how to define “passive” and “active,” it would be difficult to fully deny interest as a stimulus. Nevertheless, the relevant distinction is between what merely happens to the introspective individual and what the introspective individual actively undertakes.\textsuperscript{74}

Although in both cases the mental state is occurring either to you alone or to you as the owner of the state, it is you who self-attribute the state and form a judgment to the effect that you are in that state. Because you are the individual executing introspection, the differences in terms of modes of awareness can be preserved. SII is a caused or a sort of involuntary or unintentional event; it is the stimulus which drives the introspective awareness. In contrast, STI is an internal activity. The two thus signify respectively intended introspection and voluntary introspection.

The sense of the “active” at stake supposes that STI (a) is not automatically or immediately caused by a mental-entity (as when your laugh has been too strongly praised and thereby called notorious, for example); and is when an introspective event may be induced (such as when it is your

\textsuperscript{73} This particular view is neutral about whether other systems may play a role. However, in Ch. V., I defend a pluralist view.

\textsuperscript{74} It seems impossible to deny interest as a plausible stimulus; however, the kind of mental states involved in introspection; in particular, in the present discussion have been previously established in my Ch. I.
conscious mental state happening and you are introspectively aware of it; (b) is less limited to a specific mental state being introspectively accessed (such as the aforementioned state in (a), and (c) implies a deeper involvement from the introspective individual to select an experience or a qualitative aspect of it, to complete certain information and to furnish the experiences. So just the refining of the senses of "active" and "passive" means it is possible to elude some possible difficulties.

Some people might object that there is no determining distinction between modes of introspection, since a stimulus itself is also capable of affecting your self-generated introspection and in a way that the other sense causally works as well—meaning that in both cases an active faculty or process is in play. Alternatively, SII can produce effects in STI just as STI can produce effects in SII. I do not deny the possibility that SII and STI can mutually affect each other, such as when introspection is being initiated by a stimulus and then the output or judgment is demanding further introspection. This process generates introspection about a subsequent state. However, stimuli-induced introspection is uncontrolled and typically involves specific mental states that pop-up. Self-triggered introspection is relatively controlled and typically involves the mental states that we select.

It is not easy to reach consensus as to when STI does not participate in SII and if STI is causally inert. I accept this difficulty as it is also difficult to determine the precise boundary of "passive" or "uncontrolled" as applied to a cognitive process such as introspection—even when it is stimuli-induced or causally dependent on the target state. However, SII mainly comprises a causal relation. A mental state is causally connected to introspection if it suitably causes introspective awareness and successfully generates an introspective judgment. Alternatively, merely considering the difference between expressing reactions vs. expressing actions can help to address my point.

To determine whether or not my taxonomy is accurate is an open issue for empirical psychology and neuroscience. However, let me suggest that my qualified notion of introspection can be a tool for analyzing specific introspective cases, and can be useful for some psychological approaches in determining the underlying mechanisms of introspection—e.g., the intentional control or cognitive system.

---

Information-processing can be conceptually distinguished between SII and STI. A mechanism of selective introspection is critical to the story of classifying our conscious states, but it cannot be adequately addressed here (see my Ch. V).
To assess if I am on the right track, however, it may be worthwhile to allude to investigations on voluntary control in different mechanisms.

Empirical evidence suggests the identification of brain areas dedicated to voluntary control, and show the effects of volitions and the brain differences between voluntary swallowing and spontaneous swallowing—voluntary effort is required, for example, when individuals try to swallow a big pill (Kern, Jaradeh et. al., 2001). Additional distinctions between voluntary and involuntary brain events seem to entail different neuronal bases. Voluntary smiles are different from unintended smiles in terms of cortical activity (Iwase et. al. 2002). Moreover, research on voluntary control also helps to illustrate how those two types of control may work in tandem to execute actions (Baars, Banks and Newman, 2003; Baars, 2003), and may provide incentives for empirical researchers to investigate introspective cases.

Although a clear distinction between voluntary and involuntary control is not uncontroversial, it still requires further investigation. The relevant issue is whether some scientific findings can apply to introspection. Perhaps certain brain-imagining techniques can shed light on what areas of the brain activate when an individual engages in SII. Then we can observe whether there is a difference in terms of engaging in SIT when individuals deal with their mental states. Alternatively, we might want to find out how to distinguish between an introspection of our own states with intentional control and an introspection of our states once they have arisen. Examining these kinds of events might also lend further support to the taxonomy that I suggest here.

7. Conclusion

I have presented two leading philosophical views of introspection: the inner-sense model and the acquaintance model. I have pointed out the central problems of these views and have cleared up the sense of the terms involved.

I have shown that the definitions of introspection that both ISMI and AMI offer are too narrow and leave out important features of introspection. Instead of claiming that introspection is either causal or direct, or that there are two definitions of introspection in dispute, I claim instead that there are two different functional levels of introspection according to how this cognitive process shifts direction upon modes of introspective awareness.
As the notion of introspection gets qualified, we are able to avoid the seemingly inescapable dichotomy between the inner-sense model and the acquaintance model by resolving a dispute between these views. I have claimed that each model better captures certain elements of particular kinds of introspection or more precise characterizations of introspection which explain the relations between introspective awareness and its target mental states in different functional levels—that is, what would constitute a causal role and a non-causal role.

In searching for a definition of introspection compatible with both positions, I have argued for the possibility of the ISMI and AMI complementing, not excluding, each other. Both approaches to introspection can coexist and assist each other, and some limits of the former can be fulfilled from the latter and vice versa.

I have offered a new taxonomy: stimuli-induced introspection—or a mode of awareness that includes causal interactions; and self-triggered introspection—or a mode of awareness that includes selective and voluntary/intentional control. This taxonomy is an expansion that aims at an understanding of the nature of introspection. It is also a response to perceptual objections, and shows that introspection is a non-inferential process.

The upshot of adopting my proposal would bring about an anti-reductionist position—i.e., the psychological thesis that introspection cannot be reduced to other mechanisms, properties or entities. At the same time it would preserve the peculiarity of introspection—i.e., the epistemic thesis that introspection is a distinctive way of knowing about our minds; for there is a sharp contrast between the way in which we know our minds from other ways of knowing about the world.

Also, this view may prove useful to other theories of introspection and to researchers with a range of theoretical and empirical orientations as well. Although I have not done the experimental work that my proposal would require for its corroboration, nor does existing empirical research yet addresses these issues, my theoretical considerations would lend preliminary support.

Finally, insofar as I know, contemporary discussion has not seen the nuances that I present here. I have foreseen some clues that can be useful for developing a new model of introspection and I have motivated consideration of introspection as a process that combines and integrates several mental
phenomena (see Ch. §V). This chapter and previous chapters have already motivated the discussion to build my model of introspection.
V. A Superposition Theory of Introspection

Some philosophers claim that introspection is a single process which probes one’s mental states (single-process model); others claim that multiple cognitive processes contribute to introspection (pluralist-process model). I claim that introspection is a sui generis process that orchestrates a plural framework that combines and integrates several mental phenomena. I take advantage of some of the merits of the pluralist-process model, but I find a way to avoid reducing introspection to other processes. By taking a top-down approach to introspection, we can best understand its parts and function. To account for the phenomenon of integration, I argue that introspection is a selective, cumulative, and predictive process. I then go on to provide the grounds for constructing a model of introspection by determining variables (mental phenomena) in accordance with specific introspective cases. I suggest considering the notion of superposition to illustrate a specific way in which those variables are combined and integrated in introspection. I close by showing how a hypothetical introspective individual can determine which particular mental state is the vehicle of a particular content in an introspective episode. I contend that a practiced individual can identify changes, transitions, and boundaries between mental phenomena in specific cases of introspection.

1. Introduction and Preliminary Considerations

The philosophical study of the mind has proved highly useful in understanding the core of introspection. Contemporary philosophical theories have attempted to explain the nature of introspection in various ways.

Some philosophical theories claim that introspection is a single process, involving an “inner-sense,” a “self-scanning” process, or a “simple monitoring” mechanism of mental states (Armstrong, 1968/1993; Nichols, 2001), or that it operates as a “single mental faculty” for which interpretive and sensory access is implicit in all mental attributions to ourselves or to other individuals (Carruthers, 2009, 2011). I shall call this the single-process model.

Other philosophical theories say that introspection is a confluence of cognitive processes (Butler, 2013), or that its operation involves a variety of overlapping mechanisms (Schwitzgebel, 2012), whereby processes such as attention and memory (Goldman, 2006; Hill, 2009; Lyons, 1986) and other cognitive resources,76 such as “captioning, reintegration, and intensification” (Prinz, 2004, 2012), play an underlying role in introspection. I shall call this the pluralist-process model.

Although the competing models have been useful for understanding the nature of introspection, the initial challenge that the theorist of introspection faces is in determining whether introspection is operating as a single process or as a plurality of processes in the probing of one’s mental states.

---

76 The term “cognitive processes” here refers to perception, attention, memory, and inference. “Cognitive resources” refers to captioning, reintegration, and intensification, although these can also be identified as processes.
However, accepting any of these models implies holding a certain reductionism. Broadly construed, reductionism is the thesis that psychological theories, properties, or entities identify with physical/neural theories, properties, or entities. A type-reductionism, which is the version of reductionism specific to this discussion, holds that introspection reduces to other cognitive processes and/or properties. This view is called “reductionism of process” in the contemporary literature on introspection (Smithies and Stoljar, 2012).77

If introspection is irreducible to physical descriptions, it is unlikely that its nature can be captured by using the characteristics of other mechanisms or by providing descriptions of other cognitive processes.

If introspection were to operate as a single process, however, such as to an inner sense, it would be reducible to other processes as well, such as to perception. But then, the peculiarity of introspection would be undermined. Peculiarity is the epistemic thesis that introspection is a distinctive type of awareness or way of knowing about our minds, and that there is a sharp contrast between the way in which we know (or are aware of) our minds and what other (types of awareness or) ways of knowing about the world might be. Notice, however, that claiming that introspection is peculiar is not the same as supposing that it is the sole kind of awareness or way of learning about our minds (cf. Ch. IV., §§4.1.-4.1.1.).

Similarly, if introspection were to operate as a single process, such as what it might be within a single mental faculty (thereby using the same mechanism to access one’s mental states and other people’s mental states), it would be reducible to another cognitive process, such as an inferential process, which would consequently mean that the asymmetry between first-person and third-person perspectives would not be preserved. As a result, both the peculiarity of introspection and the privileged access would be undermined. Privileged access here refers to the epistemic thesis that introspection provides a special access to one’s mental states from a first-person perspective as compared to that

77 Contemporary philosophical theories have provided a number of arguments against various reductionist programs (e.g., Kripke, 1980; Jackson, 1982; Levine, 1983; Chalmers, 1996). Examining these responses is far beyond the scope of this project. I bypass the other forms of reductionism that may play a role here.
access being provided from a third-person perspective, even if the access being provided in the first-
person perspective is imperfect (cf. Ch. III., §§2.1).\footnote{Recall that privileged access refers to our “epistemic position vis-à-vis propositions ascribing current mental states [or processes] to [ourselves]... [This access] is favorable [or authoritative] in a way no one else’s position is” (Alston, 1971, p. 230). This results in an asymmetry of first-person and third-person, such that we do not know our minds in the same way in which we know about the minds of others. For discussions, see also Ryle, 1949; Dennett, 1991; Byrne, 2005; Carruthers, 2009, 2011.}

Therefore, the single-process model of introspection in the promotion of an inner-sense or a
mindreading faculty would be unpersuasive in either version.

However, although the pluralist-process model is promising, it also entails difficulties. If
introspection were to operate as a set of cognitive processes, or if introspection were derivable from,
conjoined with, or overlapped with other mechanisms and additional cognitive resources, it could easily
be reduced to other cognitive processes or a conflation of them. As a result, the peculiarity of
introspection would also be undermined. Therefore, the pluralist-process model would be unpersuasive,
too.

How, then, does introspection operate? Philosophical theorizing and empirical research on
mental functioning provide some hints that suggest that introspection may appeal to other cognitive
processes (or systems) and cognitive resources (or sub-systems) in order to operate (cf. Baars, 1988,
1997; Jack and Roepstorff, 2003; Menary, 2007, 2010). Hence, introspection very likely entails a pluralist
process in some way. If this is the case, the theorist of introspection is left with a puzzle: how to account
for a pluralist-process model while preserving an anti-reductionist position

Although appealing to a plural-process view that preserves an anti-reductionist position seems
difficult (let us call it the functional problem), I believe that this can succeed by construing introspection as
a top-down strategy of information-processing to analyze its parts. This cognitive strategy enables us to
work out the composition of introspection while helping us determine its function.

In order to understand how our introspective awareness works, I here claim that our introspective
awareness or introspection is a sui generis process which orchestrates a plural framework that combines
and integrates several mental phenomena. By mental phenomena I mean any mental state, cognitive
process, or resource that can occur while an individual introspectively accesses her mind. I will argue that
introspection is a selective, cumulative, and predictive process. I will then go on to provide some grounds
for constructing a new model of introspection, such as by determining variables (mental phenomena or mental features) in accordance with specific introspective cases. I will suggest considering the notion of superposition to illustrate how those variables are combined and integrated in introspection, or to show how they are functionally coordinated, while preserving the distinctions among them (a solution of the functional problem).

Taking some of the results of previous chapters, I will then propose a theoretical model of introspection and examine aspects of the nature of introspection that have been neglected in the contemporary philosophical literature. I will call this model the superposition theory of introspection. My goal here is to suggest that (what I call) “cognitive superposition” enables the elaboration of a new account of how introspection works or what the way is in which we are aware of our mental life and are able to arrive at self-attributions. Thus, I aim to show that we can determine what kind of particular ongoing mental state is the vehicle of a particular content when diverse mental phenomena occur during an introspective episode. Most importantly, I contend that a practiced introspective individual can identify changes, transitions, and boundaries between mental states in specific introspective cases.

My view builds on extant pluralist views and also extends them by advancing other proposals specific to introspection. My proposals also address relevant gaps in existing models of introspection and may fit in any theory of introspection.⁷⁹

The rest of the chapter is divided into three main sections: §2 discusses the basics of the single-process model; §3 discusses the basics of the plural-process model and highlights some of the merits of a particular pluralist view. §4 offers a justification for building a new model of introspection and sets forth its central characteristics. The chapter closes by demonstrating this model in action.

2. The Single-Process Model of Introspection

The single-process model of introspection faces difficulties on different grounds. I have previously disputed a couple of its versions: the mindreading account of introspection—aka, a single mental

⁷⁹ Although introspection could be conceived of as a type of consciousness, I remain neutral to theories of consciousness. The model of introspection that I develop, however, may be compatible with theories of consciousness: first-order (Block, 2007a; 2007b; Dretske, 1995), higher-order (Lycan, 1996; Rosenthal, 2005), global workspace (Baars, 1988; Dehaene & Changeux, 2011), and attention-based theories (Prinz, 2012).
faculty—with regard to both the unreliability of introspection and the symmetry between first-person and third-person access (cf. Ch. III., §§2.1); and the inner-sense account with regard to reductionism and the other constraints that figure in the adoption of such a view (cf. Ch. IV., §§4.1).

In this section, I claim that the single-process model of either version—that is, of the inner-sense one or the single mind-reading faculty one—falls short as an accurate account of the nature of introspection. Before making the case for this claim, it will be useful to spell out some of the basics of the single-process model of introspection.

2.1. The Inner-Sense View of Introspection

Some advocates of the single-process model claim that introspection is a simple representation of a current mental state or a mechanism of self-detection operating either as a self-scanning or a simple monitoring of our mental states. The “self-scanning process” serves to distinguish between introspective awareness and the target mental state of which one is aware—e.g., our distinguishing a pain and our introspective awareness of that pain (Armstrong, 1968/1993, 1999; Nichols, 2001). In a similar vein, this “self-scanning process” may refer to a higher-order state—in particular, a third-order thought—targeting a current state (Rosenthal, 2005; Gennaro, 2012).

Armstrong, an advocate of the single-process model, also claims that introspection is an "inner-sense" (Armstrong, 1968/1993; cf. Lycan, 1996), one that operates cognitively at a low level without intervention or influence from other cognitive processes. Armstrong argues that a “self-scanning” process is not a complex process such as what one finds in deliberation; nor is introspection “inferential.” Instead, introspection works as a “simple detection” of a mental state that is occurring to oneself (Armstrong, 1993, pp. 325-7; 1999, p. 114).

Armstrong expands by writing such ideas as: “We simply become aware of some current mental content in the same sort of way that, in vision, we become aware that something or other is before us.” Moreover, “[t]he introspective mechanism, whatever it is, does no more than keep a watching brief on our own current mental contents, but without making much of a deal about it.” “Like perception, introspection is a matter of representation. But it is representation with a narrow domain: nothing but the representee’s

---

80 Lycan (1996) endorses an inner-sense view that involves attention. I bypass discussion of this issue.
own (current) mental life. It can seem rather a mystifying thing, and it would help if we could find a demystifying model for it” (Armstrong, 1999, pp. 114-16, my emphases).81

Several philosophers, such as Butler, Byrne, and Shoemaker, have reacted against the inner-sense view of introspection. They claim that one problem with the inner-sense view concerns the terms used to explain introspection and another problem concerns the specific role that the first-person method plays in such a view. Butler (2013) argues that the use of “inner-sense” or perceptual terms to explain introspection is only metaphorical, and that those terms are inadequate in explaining “what really goes on when we introspect” (p. 91). Byrne (2005) puts forward another reaction: contrary to introspection, perception does not involve a first-person method of knowing or being aware, for example, “that one believes that the cat is indoors, that one sees the cat, that one intends to put the cat out, and so on” (p. 80; cf. Shoemaker, 1996).

Although the inner-sense view correctly distinguishes the introspective process from its target mental state and recognizes that introspection is a non-inferential process, taking introspection as an inner-sense or a simple self-scanning method by which we merely reach awareness of our mental states would make it reducible to a perceptual-like process.

Notwithstanding their enthusiasm for establishing a single-process model of introspection, however, the advocates of the inner-sense view overlook the main risk of embracing a reductionist approach: it undermines the peculiarity of introspection.

2.2. The Mindreading Faculty View of Introspection

The single-process model is also endorsed by those who claim that introspection is a kind of “self-interpretation” of mental states based upon one's experiences and behavior (Carruthers, 2009) and operates as a “single mental faculty” or “mind-reading” system that underlies attributions of both our mental states and other individuals’ mental states—in particular, propositional attitudes such as decisions and judgments (Carruthers, 2011).

---

81 Although the claim that introspection and perception are similar is controversial and the claim that perception operates at a low level without influence of other processes—or states—is also controversial, I bypass discussion of these problems; see Siegel 2010, 2017.
This single mindreading faculty involves both “interpretive access” and “sensory-based access” that allow the individual to learn about attitudes through speech and behavior. However, this single faculty operates cognitively at a low level without its requiring a demand from other cognitive processes. The same mindreading faculty takes sensory data as input, interprets these data by drawing on situational facts and background information, and yields representations of mental states as output (Carruthers, 2011, pp. 19, 163, 215).  

Some philosophers have reacted against the mindreading view. Rey (2012; forthcoming) argues that the problem with this view concerns the presupposition that attitudes are frequently attributed to explain non-self-attributive behavior. “It's because he [Carruthers] thinks decisions and judgments, so construed, can be shown to diverge from what people avow that he thinks decisions and judgments cannot actually be introspected, but only inferred as part of an explanation of, e.g., motor performance and the satisfaction of goals” (p. 9).

Rey argues that the mindreading view disregards the role of inner speech in the introspection of attitudes. This may be problematic, since the internal monologue and “the sincere avowal of an attitude [can serve] as evidence of a deeper, internal, 'introspective' distinction” (Rey, 2012, forthcoming, p. 9 and personal communication). Moreover, the mindreading view tends to leave out the standard use of “attitude concepts” for claims such as “I think that ρ” (Wright, 2000). Rey adds that the first and third-person could apply attributions in the same way to claim that “I hope that ρ” merely by noticing that that person is disposed to sincerely self-attribute such a hope. However, the way “we hear speech” (Fodor, 1983), and “the phenomenology of the self and the present” (Peacocke, 2012) involve a kind of awareness of an “I” and “now” that only entails the first-person.

Contrary to the mindreading view, other philosophers have argued that we are able to introspect, rather than infer, concurrent propositional attitudes in a way in which we are not able to know about such attitudes of other people. For example, Shoemaker (1996) argues that it would be hard to accept that individuals are “self-blind” or that they do not possess the rational and conceptual abilities to introspect such states. “A self-blind creature would be one which has the conception of the various mental states,

82 Although the claim that “other-knowledge subserves self-knowledge” is closely linked to the claim that our basic conception of mental states—“folk psychology”—derives from theorizing; this topic goes beyond the present purview; see Gertler (2011).
and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way," is one way he characterizes this (pp. 30-31).83

Since the mindreading faculty requires an interpretive access that makes use of an inferential process, precisely of the sort we make about others' attitudes, and since this inferential process gets information from our sensory system, behavior, and context, introspection is easily reducible to an inference-like process. That is, the explanation of introspection is reduced to the explanation of a different process.

The mindreading view not only reduces introspection to an inferential process; it also leads to the claim that we gain introspective access to our mental states in the same way in which we gain access to other people's mental states—by drawing conclusions on the basis of observed behavior. In doing so, the advocate of the single-process model overlooks the risks of embracing a reductionist approach. This undermines, respectively, the peculiarity of introspection and the privileged access.

Other advocates of the single-process model account for a "self-monitoring" mechanism and present some errors of the mindreading faculty view. The self-monitoring mechanism view distinguishes between self-monitoring conscious states and self-monitoring non-conscious states (such as propositional attitudes), and also distinguishes self-monitoring from the mindreading faculty (which deals exclusively with the mental states of other people). Thus it predicts a "double dissociation" between metacognition and mindreading (Nichols and Stich, 2003).

The advocates of the self-monitoring view examine cases to show that even if the self-monitoring mechanism breaks down, the inferential process of the other's mind remains intact. They also examine cases to show that even if the inferential process might be impaired, the self-monitoring mechanism still continues operating normally. That is, they provide evidence of instances of people who can attribute mental states to others successfully, but who have difficulty in attributing mental states to themselves—such as in individuals with the passivity symptom of schizophrenia. Also, they provide evidence of

83 Contrary to Shoemaker, Gopnik (1993) claims that we become aware of our mental states by applying a theory to ourselves just as we do to other people (the "theory-theory"). "Intentionality" names a theoretical construct that we invent early in our lives to explain behavior and actions about ourselves and other people.
instances of people who maintain reliable access to their own mental states while losing their capacity to attribute such states to others—such as in individuals with autism. The data suggest that the inferential process and the self-monitoring process are not only different and separable processes, but that introspection also does not require the intervention of other processes (Nichols, 2001; Nichols and Stich, 2003).

In fact, cognitive scientists have shown that equating introspection with a single process is inaccurate. Indeed, empirical evidence suggests the contrary. Introspection works on many different kinds of states and processes, and it works in coordination with several cognitive processes and resources. In the context of preserving introspection as a plausible scientific method for understanding various aspects of the human mind, research on emotions, beliefs, desires, various aspects of our phenomenology, problem-solving performance, and assistance with the medical and clinical treatment of various diseases and disorders such as synesthesia, autism, and so on, show that introspection does not seem to operate as a single process (cf. Jäkel and Schreiber, 2013; Cytowic, 2003; Zahavi and Parnas, 2003).

In support of this view, others argue in favor of a pluralist account of mental function that may apply to introspection as well. “The full spectrum of cognitive abilities is relevant...memory, attention, imagination, learning, decision-making, problem-solving, conceptualization, language, and so on. There are many different kinds of cognition, enabling a broad range of cognitive capacities involved in human life” (Butler, 2013, p. 75). “It is quite natural to be pluralistic about cognitive processes and vehicles; as such, there is no single genuine ‘cognitive kind’” (Menary, 2007, p. 13; 2010).

So far, we have seen that the single-process model tends to use reductive arguments to address the nature of introspection. We are in a position to claim that introspection is a sui generis process; and that no matter how closely we look at it, it is irreducible to a simpler process, whether akin to the perceptual or inferential.

Since the explanation of the inner-sense and the single mindreading faculty views does not suffice for the explanation of introspection, the single-process model is unpersuasive.
3. The Pluralist-Process Model of Introspection

In this section, I claim that even though the pluralist-process model is more compelling, it falls short as an accurate explanation of the nature of introspection. We will see that different versions of the pluralist-process model require a different treatment. According to Prinz and Schwitzgebel, pluralism means that introspection is “varied”—i.e., different processes can occur on different occasions. In contrast, according to Goldman, pluralism means that introspection is rather “complex”—i.e., plural processes occur in every given case. Although I accept the complexity of introspection *qua* process, I do not think that plural processes occur in every given case (see §4.3. on *variables* of introspection).

An advocate of the pluralist-process model claims that introspection always works within processes such as perception or attention (this is the “introspective self-attribution” view; see Goldman, 2006). But this view also reduces introspection to other cognitive processes.

Another advocate of this model defines introspection as a mere confluence or overlapping of processes and resources (this is the “spaghetti” view; see Schwitzgebel, 2012). But this reduces introspection to a conflation of cognitive processes.

Finally, another advocate of this model defines introspection as a collection of multiple things and claims that there are different “species” of introspection (the “fractionation” view; Prinz, 2004). But this view seems to reduce introspection to multiple mechanisms of access.

Before presenting the case for each of these claims, it is useful to spell out some of the basics of the pluralist-process model of introspection.

3.1. The Introspective Self-Attribution View

In opposition to the single-process model, there are also philosophers who argue in favor of the pluralist-process model of introspection. In this model, introspection operates at a higher level cognitively and with the intervention or influence of specific cognitive processes. So it is complex. In particular, introspection works within processes such as perception and attention (Goldman, 2006), or attention and memory (Hill, 2009). The former processes working together with some cognitive resources such as “recognition,” “redeployment,” and “translation” (Goldman 2006) are central to the formation of introspective judgments or self-attributions.
Let us focus on the Goldman's view only—aka, the *self-attribution view*—to show how it works. After, let us comment roughly on Hill’s view.

According to the self-attribution view, introspection is a “quasi-perceptual” process in that it classifies its target mental states in a manner similar to that in which perception classifies its objects. To do so, it involves attention. Goldman (2006) argues that “[i]ntrospection is perception-like, however, in the way it recognizes or classifies the target state in terms of its general category and in terms of such characteristics as strength or intensity” (p. 247). He adds that “the input to visual processing” is central to introspection and that “the ‘organ’ of introspection is attention, the orientation of which puts a subject in an appropriate relation to a targeted state” or “attention seems to act like an orienting organ in introspection, analogous to the shift of eye gaze or the sniffing of the nose” (p. 244).84

Goldman presumes that the introspective process is capable of performing some information-processing operations that are analogous to perception. Three different cognitive operations, then, subserve introspective judgments or play a role in introspective attribution: “recognition,” “redeployment,” and “translation.”

“Recognition is used in typing the target state” that comprises beliefs, emotions, and bodily sensations, and in “classifying the target state in terms of supplementary features like strength or intensity.”

“Redeployment is used…to represent the particular content of a mental state by…copying or replicating the content of the introspected mental state into the content of the introspective judgment.”

“Translation [is used to convert] from one mental code to another” or to change “the target state's content in beliefs when that content is represented in a visual format” [in order] “to judge if they are to be introspected. Visual representations [for example] have a different mental code than beliefs; hence, some cognitive work is required to translate the fine-grained detail of visual experiences into mental contents that can be judged introspectively” (Goldman, 2006, pp. 253-55, my emphases).

In response to the self-attribution view, one may say that introspection and perception, or introspection and attention, operate “metaphorically” in the same way (Butler, 2013). It is difficult to

---

84 For differences between "perceptual attention" and "introspective attention," see Shoemaker, 1996, Ch. 10. In my chapter I, attention is considered an epistemic condition of introspective access.
determine their specific role in that analogy and to discern when exactly introspection operates, and when it is either perception or attention.

Again, "although we commonly understand introspection in perceptual terms (due to the utility of perceptual metaphors in getting a handle on our understanding of ourselves), this is not an adequate account of what really goes on when we introspect. Similarly, one may argue that our common-sense, metaphor-driven understanding of attention does not accurately map onto what is really going on in our minds." So we need a different account of introspection (Butler, 2013, p. 91).

Some theories of mind consider that separating cognitive processes and vehicles should be the desideratum for any theory that aims at explaining introspection:

[T]here are functional and physiological reasons for separating attention processes from perception processes, despite the fact that they work together in various ways to produce our mental lives as we experience them. Most notably, there is pertinent neurological evidence that perception and attention or, more specifically, primary sensory processing and secondary attentional processing involve identifiably different areas of the brain…the posterior regions of the primate neocortex are devoted to processing sensory/perceptual data, while the frontal regions seem to involve the more supervisory, monitor-like cognitive domains associated with attention (Fuster, 2002; Picton et al., 2002). If attention and perception occur in noticeably different areas of the brain, as these data suggest, we…should not regard the explanation of introspection through attention mechanisms as supporting the inner…model. These are very different kinds of processes and should not be confused with one another (Butler, 2013, p. 39).

Unquestionably, there are important parallels in the operation of introspection and attention. Both seem to involve, for example, the processing and organizing of information from mental states and then directing it to particular contents, and the applying of conceptual resources to the mental states. Nonetheless, introspection and attention are different processes:

"Attention…can be directed toward either internal or external stimuli (or perhaps simultaneously to both), depending upon what aspects of the attended object are being attended to." For instance, in attending to a visual experience, "I can attend to my experience of the visual field by reflecting upon the
fact that I am having the experience, or I can turn my attention outward toward that which I am seeing…” Moreover, I can pay attention to an experience and also to what it represents, whereas in introspection the content is necessarily attached to the target mental state—i.e., we do not introspect the representation itself, but the state of pain (Butler, 2013, p. 90-3, my emphases).

The self-attribution view seems to entail similar, if not the same, problems of the single-process model about the reductionism of introspection to other processes. The explanation of introspection would be reduced to the explanation of other processes as well. Even if the intention of the advocate of the self-attribution view were, instead, to simplify the explanation of introspection, the self-attribution view might be oversimplifying the introspective process itself. This does not sound compelling from a pluralist perspective.

There is, it is worth pointing out, the figure of the advocate of the plural-process model of introspection who does not consider introspection to be a perceptual process. In fact, the plural-process advocate recognizes that there are important differences between these two processes. Introspection is a process that generates judgments about our mental states rather than being a kind of perceptual awareness about the content of those target states. The plural-process advocate considers attention and memory fundamental tools for the functioning of introspection.

The plural-process advocate of this secondary view claims that introspection is not a single process since it is unlikely that a mere self-scanning event operates for all of the target mental states. Hill claims that introspection varies depending on the type of target mental state, and that introspection about standing or enduring beliefs and desires involves a complex procedure for searching “vast and heterogeneous” long-term memory stores (Hill, 1991, 2009). He also claims that attention is important in introspection since it modifies, shapes, and even creates the introspected target mental state. In this respect, the plural-process advocate of this view argues that introspection differs from perception, since the latter does not modify the object perceived (Hill, 1991).

Although this view would not seem prima facie to reduce introspection to other processes, a couple of additional worries become salient. It is relevant to consider that, as opposed to retrospection (which refers to a reflection of strictly past events and experiences), introspection focuses on concurrent states and processes as they happen in our mind or occur in the stream of consciousness. This enables
us to describe and report our thoughts and experiences.\footnote{`Stream of consciousness’ roughly refers to numerous mental states and processes—and associations—that go through the mind as an individual’s internal monologue. It also refers to a narrative mode that involves thought processes from a first-person perspective. See James, 1980/1981.} This view of introspection does not imply that some dose of \textit{short-term} memory is unnecessary for introspecting (cf. Ch. I-Ch. II, §§4-4.1). I have discussed how introspection only alludes to \textit{occurring} beliefs and not to standing beliefs, which are among the objects of retrospection (cf. Ch. III §§2-2.1).

Alternatively, other philosophers endorse the claim that introspection \textit{overlaps} with diverse processes: for Butler (2013), “introspection is a multi-faceted phenomenon that cannot be limited to a single cognitive mechanism or epistemic characterization.” That is to say, it “is not a single faculty of the mind, as is commonly thought, but rather ["a pluralistic confluence of processes" or] a heterogeneous collection of different kinds of first-person experiences and cognitive processes” (pp. 1, 161).

According to Butler (2013), introspection makes use of three processes in its operation: \textit{representation} of “our own [target] mental states” or objects of introspective awareness; \textit{conceptualization}, which allows “us to place the various phenomena we experience in the coherent theoretical networks that make sense of them”; and \textit{attention}, which “enables us to guide these cognitive resources towards various phenomena such that we can actively reflect upon and think of our mental lives” (p. 162).

Although this view would not seem \textit{prima facie} to reduce introspection to other processes, one might claim that perception also makes use of these processes in order to operate. Hence, it would be reasonable to claim that representation is as characteristic of introspection as it would be of any form of awareness, whereas conceptualization and attention are epistemic conditions of introspective access (cf. Ch. I. §§1-3; Ch. II).

Even if attention and introspection are interrelated in some of the mentioned ways, the advocate of this pluralist version fails to provide an explanation of exactly how these two cognitive processes interact or operate together and how they can preserve their respective differences or be identified as separable mental phenomena.
3.2. The Spaghetti View of Introspection

An advocate of the plural-process model claims that "[i]ntrospection is not a single process but a plurality of processes" where "there is no shaped line to determine a genuine introspective process isolated from related, adjoining, and overlapping processes." Introspection is a mere "cognitive confluence of crazy spaghetti"—henceforth, the spaghetti view (Schwitzgebel, 2012, p. 29; see figure 1).  

I doubt that we can draw sharp lines through this knotty snarl, cleanly isolating some genuinely introspective process from related, adjoining, and overlapping processes. What we have, or seem to have, is a cognitive confluence of crazy spaghetti, with aspects of self-detection, self-shaping, self-fulfillment, spontaneous expression, priming and association, categorical assumptions, outward perception, memory, inference, hypothesis testing, bodily activity, and who only knows what else, all feeding into our judgments about current states of mind. To attempt to isolate a piece of this confluence as the introspective process – the one true introspective process, though influenced by, interfered with, supported by, launched or halted by, all the others – is like trying to find the one way in which a person makes her parenting decisions, the one cognitive process behind writing a philosophical essay, [and so on] (Schwitzgebel, 2010, p. 17).

Although not stated explicitly, it is likely that the spaghetti view might be inspired by the fractionation view (proposed by Prinz, 2004; see my §3.3)—in particular in the pluralist claim that introspection is "a jumble of different things" (p. 46). Nonetheless, contrary to the former view outlining a "cognitive confluence of crazy spaghetti", the fractionation view advances an account of how some particular cognitive processes and resources cooperate within introspection, and thus it provides a plausible solution to the problem of conflation. I shall turn to this below in §3.3.

Schwitzgebel’s spaghetti view (2010) stresses the claim that introspection consists in the operation of many different processes, including perception, memory, inference, and several cognitive aspects as well. However, while it is very likely that some of those cognitive processes take place in introspection, according to the spaghetti view we cannot separate them. For example, introspective judgments and perceptual judgments are deemed indistinguishable: "Judgments about sensory

---

86 I have disputed Schwitzgebel’s view with regard to the unreliability of introspection in Ch. II.
experience can easily collapse into judgments about the outside world” (p. 7). Also, we cannot separate processes such as *perception* and *introspection*: “The processes by which I see are part of, or overlap with, the processes that shape my [introspective] judgments about the resulting visual experience” (p. 8).

Figure 1 illustrates the “boxology of introspection” or the “spaghetti” view offered by Schwitzgebel (2012, p. 40).  

Some philosophers have also reacted explicitly against the spaghetti view. They believe that even if we accept that “pluralism consists in the claim that... [introspection] is multiply realized by the operation of many different psychological processes at the personal level, including perception, inference, and so on,” the spaghetti view that construes introspection as a mere overlapping of different processes is meant to undermine the peculiarity of introspection (Gertler 2012; Smithies and Stoljar, 2012, pp. 7-8; see also Sosa, 2012; Zimmerman, 2012).

The relevant point is to see that the “spaghetti view” is particularly troubling in the light of any philosophical theorizing or psychological account of introspection. Even if we grant that introspection operates in pluralistic terms, as a confluence of different cognitive resources and processes, this view does not provide an accurate picture of what introspection is. It is unlikely that introspective awareness *qua* cognitive process works in such a conflated, convoluted or indistinguishable way.

If introspection results in an overlapping of several processes, the advocate of the plural-process model must explain what such an overlapping of processes amounts to. While it is certainly hard to determine how introspection operates given that diverse processes seem to not only hold together (and perhaps crowd up), but also occur simultaneously, it is precisely the task of the theorist of introspection to shed light on this issue and untangle the processes.

---

87 My model offers an alternative to this representation at §4.1.
It is one thing to claim that introspection involves other processes and resources in order to operate, or that a diversity of processes and resources cooperates in introspection for input (either of which is sound). But it’s another to claim that introspection is a “confluence of crazy spaghetti.” The former claim would lead us to investigate how those processes and resources are integrated in introspection, whereas the latter would indicate a halt.

So far, we can conclude that since the spaghetti view reduces introspection to a conflation of processes and resources, the pluralist-process model is unpersuasive.

3.3. The Fractionation View of Introspection

In opposition to the single-process model, there is another philosopher who advocates for the pluralist-process model. The thinking goes that anyone assuming that introspection is a “single phenomenon may be making a mistake… [since] introspection is a jumble of different things.” Rather, there are different “species” of introspection or different independent routes to introspection. Introspection “splinters into several different species, involving different underlying mechanisms.” This is what Prinz calls the *fractionation view* (Prinz, 2004, pp. 40, 46).

In fact, Prinz claims that “[t]here is much to be gained by distinguishing different species of introspection. If dissociations are possible, then each species can be investigated independently, and a distinct mechanism might be identified” (p. 48). To clarify, “species of introspection are not unified by a common mechanism for accessing information.” Instead, these “recruit different mechanisms of access” which share a common feature: accessing mental states, without entailing an “underlying unity at the level of mechanism” (p. 53).

Prinz’s fractionation view presumably considers introspection as something coming in different species; this view seems to involve multiple processes (different processes on different occasions), depending on both the target mental state at issue (whether it is a conscious experience or an attitude) and the particular mechanism of access to those mental states (whether it is controlled attention, a passive noticing, verbal “captioning,” or a labeling of experiences). Prinz also stresses that introspection often intensifies or otherwise modifies the target experience. In such cases, introspective “access” is access only in an attenuated sense. But let us set aside this point and focus on the view.
The fractionation view offers a taxonomy of species of introspection, which it spells out as follows: First, *captioning* consists in a “verbal” or a “non-verbal access” to the contents of conscious states. This allows us to “describe the process of narrating” those conscious states. We use a sort of “labeling [of] an experience” in order to be aware of “what … [we are] experiencing.” For instance, “[i]n an individual with intact hearing, this narration is experienced as subvocal speech…a stream of auditory images of words—[that]…become conscious in just the same way that perceived sounds become conscious” (pp. 40-41, 50). Moreover, captioning carries relevant conceptual distinctions. In verbal access, there is “psychological captioning,” which refers to our use of “a first person pronoun,” and there is “mere captioning,” which refers to when we do *not* use it. For example, labels such as “there is a tomato” or “this smell is good” are different from utterances like “I am smelling the tomato or I am seeing red.” Additionally, captioning may generate “different kinds of self-reports [which] carry [in turn] different conceptual requirements,” like elaboration or interpretation (pp. 41-42).^{88}

Second, *reintegration* is “a matter of broadcasting mental states throughout the mind” without necessarily implying a certain access to them. It rather suggests the use of working memory [where we store information temporarily for active use] to connect our target mental states to information coming from our sensory modalities and executive faculties. Reintegration principally involves the intervention of working memory “to encode a current [conscious] state” (pp. 44, 53).

Third, *intensification* consists in “a matter of [change or] increasing vividness or detail” of a particular conscious state, which allows “more information to propagate forward.” This process is considered “as an increase in attention resulting [in turn] in an increase in the portion of a perceptual stream becoming available to working memory.” Intensification principally carries the intervention of attention, which can modify the introspective target’s mental state (pp. 44-45, 50).

The fractionation view defines introspection as a “jumble of different things,” all of which might lead to a conflation of cognitive processes. The view also claims that there are “species of introspection.” This seems to overlook the point that introspection may be an underlying mechanism *coordinating* different cognitive processes. However, the fractionation view provides hints to distinguish a range of

^{88} For a discussion of the notions of ‘I,’ ‘the self,’ ‘agency,’ and ‘ownership’ in this context, see Prinz 2012, Ch. 7.
different cognitive processes as they cooperate within introspection. In this view, introspection would involve “captioning,” “reintegration,” and “intensification,” all of which can occur independently. And when they do, each of them may be called introspection. Yet it would not be accurate to call this a species of introspection.⁸⁹

Therefore, this pluralist approach may show us a plausible way to understand how certain cognitive processes and resources play a role in introspection. The mere categorization of cognitive resources is relevant: captioning uses conceptualization; reintegration uses memory; intensification uses attention.

We have seen that other pluralist views reduce introspection to different processes and perhaps a conflation of them. However, the fractionation view may shed light on the operation of introspection, making the view compelling for that reason. At the very least, the fractionation view may show how attention and conceptualization play a role in introspection, and may locate how short memory cooperates.

Additionally, the fractionation view may provide some hints for construing introspection as a top-down cognitive strategy. However, this approach still requires some refinements. For it leaves unexplained what specific kind of relation introspection entails with those cognitive processes and resources. This view, then, does not suffice to explain how other cognitive processes and resources are integrated in the introspective process.

As I have said, a plural-process model may be promising if it addresses the difficulty of reductionism. The question now is whether we can make use of the fractionation view, in particular, while preserving an anti-reductionist position. This is the challenge for the theorist of introspection.

To illustrate the core of the previous section (§3), see the following table:

<table>
<thead>
<tr>
<th>Single-process Model</th>
<th>Reductionism of introspection to other cognitive processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Inner-Sense View</td>
<td>Yes, it is reduced to perception.</td>
</tr>
<tr>
<td>• Mindreading View</td>
<td>Yes, it is reduced to inference of a specific kind (theory of mind).</td>
</tr>
</tbody>
</table>

⁸⁹ Prinz (2004) considers that we can introspectively access memories. However, I claim that introspection has access to limited mental states (cf. Ch. I). I set aside these “species” of introspection or what he calls “episodic retrieval,” since I consider it to be specifically related to retrospection instead of introspection.
Plural-process Model
- Self-attribution View Yes, it is reduced to perception and/or attention.
- Spaghetti View Yes, it is reduced to a conflation of processes.
- Fractionation View Yes, it is reduced to multiple mechanisms of access.

To illustrate the core of the next section (§4), see the following table:

- Alternative View

<table>
<thead>
<tr>
<th>Anti-reductionism of processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introspection is irreducible to a perception, inference, or attention-like process, and to a conflation of diverse processes and/or multiple mechanisms.</td>
</tr>
<tr>
<td>Introspection works on many different kinds of mental states and processes, and it works in coordination with several cognitive processes and resources.</td>
</tr>
</tbody>
</table>

4. A Superposition Theory of Introspection

This section is divided into four subsections that argue for my own proposals. §4.1. begins by construing introspection as a top-down strategy of information-processing and offers an alternative view. §4.2. describes how introspection can be functionally coordinated with other cognitive processes. §4.3. identifies specific variables (mental phenomena) that occur in introspection and draws on how these variables can be combined and integrated in introspection (as opposed to the jumble and spaghetti views). §4.4. accounts for a function—i.e., a “superposition”—which would be a specific kind of integration and combination of variables that involves an additive effect and provides a proportional result. This function illustrates how diverse mental phenomena occur in particular introspective cases and how mental phenomena can be distinguished in specific introspective episodes. I shall call this cognitive superposition.

Before proceeding, it is worth mentioning that an introspective episode is thought of as being composed of a beginning, middle, and end. An introspective episode can begin by a self-triggered introspection and end up with a verbal report, or it can begin by a stimuli-induced introspection and end up solely in a silent judgment. Even if the episode of introspection is relatively short, and only encompasses one’s current, ongoing, and recently past mental states, it is distinctive of this process that its duration is susceptible to further extension if the process continues as it might (or might not).
Scope of introspective episode

Begin

- Modes of awareness: Stimuli-induced introspection or self-triggered introspection

Middle

- Process of awareness: self-attributions or introspective judgments

End

- Deliverances: Judgments (silence) or verbal reports (or listening to one's verbal reports)

4.1. Anti-Reductionism of Introspection to other Mental Processes

In what follows, I will provide the contours of what can be considered a defense of a certain “pluralist” anti-reductionist approach. I will leave some of the details for future work. However, I hope to offer the foundations for a theoretical view that explains what introspection is and/or how it works—I will call it the superposition view.

As I have said, if the theorist of introspection advocates a plural-process approach, she needs to preserve an anti-reductionist position—namely, that introspection is irreducible to other cognitive processes or to their properties (reductionism of processes).

An anti-reductionist position maintains that introspection entails a process of awareness and learning about our mental lives and, while being ineluctably distinct from other cognitive processes, operates in tandem with other processes and makes use of a number of mental phenomena without losing track of them.

I claim that a proper superposition view of introspection can address the challenge of reductionism by construing introspection as a top-down strategy of information-processing that provides a picture of how the system works and analyzes its parts. This “cognitive” strategy is goal-directed and enables the working out of the composition of introspection while helping us to determine its function.

In models inspired by the literature on computer science, it is commonly said that information flows in two directions: bottom-up and top-down. In its simplest terms, a bottom-up direction refers to any flow of information from the earliest stages of sensory processes forward, while a top-down direction refers to any flow of information in the other direction that comes from the latest stages of incoming information. A definition implicit in the literature on modularity draws a cognition/perception distinction and
defines top-down as a perceptual processing that is modulated by cognition. Another definition commonly related in the literature on attention defines top-down as a sensory processing that is, in some way, volitional or under voluntary control. Space prevents us from discussing the different definitions and connotations of the terms “bottom-up” and “top-down” and their problems (cf. Shea, 2013).90

Through this strategy, we can appreciate (1) that the operation of introspection is distinct from that of other cognitive processes, (2) that several mental phenomena simultaneously occur in self-probing a target mental state, and (3) that these phenomena can be distinguishable, irrespective of individual brain structures, individual cognitive abilities, background information, and capacity to follow tasks, attend to certain things, and structure information in a particular way.91

More specifically, I aim to argue that introspection is a cognitive process that makes use of a top-down approach to distinguish itself from other processes, such as in memory or attention, or in the dissection of mental states in terms of both character and content, and in the distinguishing of other mental phenomena that occur in introspection.

Before proceeding to explore how introspection can do this, it is necessary to make some adjustments to the fractionation view, perhaps even to the point of redefining the categorization that this view offers.

In furtherance of both improving the plural-process view so as to preserve an anti-reductionist position, and to propose an alternative view that admits of a pluralist and anti-reductionist position, I suggest that:

A. Rather than claiming that introspection is "a confluence of crazy spaghetti" and/or "a jumble of different things"—as Schwitzgebel (2010, 2012) and Prinz (2004), respectively, suggest—it should be agreed that several distinguishable cognitive processes operate in tandem with introspection. To avoid conflating them, we need to explain how introspection can be functionally coordinated with other processes. Accordingly, I argue that introspection is a selective, cumulative, and predictive process. This proposal will allow us to appreciate how other processes

---

90 Bottom-up is also called stimulus-driven processing, while top-down is called knowledge-driven processing. It is not my intention to connect these strategies of information-processing to stimuli-induced introspection and self-triggered introspection (cf. Ch. IV), even if such a connection may work.

91 Personal communication with Gabriel Reyes (see Reyes & Sackur, 2014; Fleming et al. 2014).
are cognitively integrated in introspection. That is, this suggestion explains the phenomenon of integration (§4.2), which can be summarized as follows:

**Avoiding conflation of mental processes by identifying introspection as:**

- Selective process
- Cumulative process
- Predictive process

**Processes integrated in introspection:**

- Attention
- Memory
- Inference

B. Rather than claiming that there are *species* of introspection—as Prinz (2004) suggests—it should be agreed that there are *variables* of introspection which trigger certain processes. Thus “captioning,” “reintegration,” and “intensification” can be *variables* of introspection that use conceptualization, memory, and attention. Also, it should be agreed that these variables can be sub-categorized, respectively, in terms of *scope* (possession/use of conceptual resources to label mental states), *level* (accuracy in encoding information) and *degree* (intensity or richness of detected experience). This suggests that there are variables of introspection that are cognitively integrated within the process (§4.3).

Recall that variables are mental features or mental phenomena—meaning any mental state or cognitive process—that occur while an individual introspectively accesses her mind and which can affect the introspective result. Or else we can consider variables as internal noise or interference that can influence or affect the introspective process. Although several mental phenomena simultaneously occur in self-probing a target mental state, these phenomena can be distinguishable—or so I will argue.

---

**Mental phenomena** that are integrated within introspection

- **Variables cognitively integrated:**
  - Captioning” (verbal/non-verbal)
  - Reintegration
  - Intensification

- **Processes triggered/used by variables:**
  - Conceptualization
  - Memory
  - Attention

- **Scope of possession/use**
  - Level of accuracy
  - Degree of richness
C. Rather than considering the cognitive processes and resources involved in introspection as indistinguishable, we have reasons for an alternative account of information-processing that can postulate a “superposition.” Superposition is not a mere integration of variables, but a particular kind of integration and combination of variables that brings about an additive effect and a proportional result. This approach provides a simplified account of the operation of introspection by alluding to its minimal elements while revealing its central function. It also offers a conceptual differentiation of the variables (mental phenomena) involved in particular introspective episodes. Additionally, it offers a plausible way to explain how the variables of introspection work or can be connected. This suggestion explains how the phenomenon of integration, specifically as a superposition, operates (§4.4).

Figure 2 illustrates how two variables are superposed, in the simplest form. Contrast this graph to the representation of the “spaghetti view” (Figure 1, §3.2.).

To show what superposition can do, I suggest that:

D. A superposition theory can explain how introspection operates. This theory accounts for the way our introspective awareness can reveal which particular mental state is the vehicle of a particular content in an introspective episode. It can also reveal changes, transitions, and boundaries between mental phenomena in specific cases of introspection (§4.4.1.).

4.2. Integration of Processes in Introspection

I turn now to showing how introspection can be functionally coordinated with other processes. I argue that introspection is a selective, cumulative, and predictive process. This view may help us to see how other cognitive processes may play a role in introspection—that is, how processes such as, respectively, attention, memory, and inference may be cognitively integrated in introspection.

92 For the rest of the section, it would be useful to bear in mind this representation. For animation of two waves, see https://commons.wikimedia.org/wiki/Category:Superposition_principle
These characteristics (selective, predictive, and cumulative) can be considered aspects of the nature of introspection or features that subserve the introspective process in order to operate. The relevant point is that these features are helpful in providing an explanation of the way introspection structures and organizes information. This means they may also help in explaining the peculiarity of introspection.

Introspection entails a process of selection, accumulation, and prediction of several mental phenomena, whereby we come to learn about our psychological lives, insofar as several conditions obtain. Specifically, these conditions include the basic conditions of introspection (mental, occurring, and first-person) and the epistemic conditions of introspective access (attention, conceptualization, situation, justification, authority of error, possession of the right cognitive capacities) (cf. Ch. I-III).

Introspection is selective in probing an occurring mental state. It comprises the retention of current or recently past mental states and the accumulation of introspected episodes. This makes it possible to anticipate and predict new psychological responses in specific introspective episodes for which the individual may or may not be able to provide an explanation. The relevant point in this respect is that the introspective individual possesses first-person access to her mind and from such access she can predict subsequent mental phenomena.\(^93\)

These characteristics may be verifiable for the most part by empirical psychology. In fact, imaging technology may soon be able to reveal which areas of the brain are active when the individual performs certain introspective tasks, provided that these features (selection, accumulation, and prediction) occur in the introspective processes. However, even independently of such confirmation, theorizing about them may help to pave the way for determining how introspection works in the first place.

\(^93\) The predictive coding hypothesis holds that the brain continually generates models based on context and information from memory to predict sensory input. Also, these models take in incoming sensory information based on previous mental states, so that expected information is processed efficiently and resources can be allocated to novel or surprising information, and can be used in updating information according to error. Although my suggestion is not intended to allude to predictive coding models, it is worthwhile mentioning that they may eventually help us to figure out how introspection works on a neurological basis (see Friston, 2005; Hawkins & Blakeslee, 2004).
(a) Selective Process

Introspection entails the selection of a mental phenomenon from a set or combination of mental phenomena. Since this process involves the focusing on one’s concurrent mental states and then the pointing out and making the choice of one mental state over the several alternate states occurring in the one’s stream of consciousness, certain shifts in the introspective process will be derivable. Therefore it is convenient to characterize this process as a selective process.

Introspection is selective as it probes a target mental state with certain content among an array of mental states with a particular qualitative character or mental attitude, thus enabling the generation of judgments about the state, either in coarse or fine-grained terms. More specifically, introspection is a selective process that combs through several mental phenomena as they happen in the mind. While ignoring other mental phenomena, it concentrates on a specific state from within a span of mental phenomena or a succession of diverse experiences and thus becomes a process of decision-making from which the character of that particular state is learned.

In this respect, recall that I have defined self-triggered introspection as an active or selective process in which the individual’s own interests or volitions initiate introspective awareness. This mode of introspective access is not caused by a stimulus—that is, no inputs or mental stimulus engender the initial introspective episode. Instead, the self-triggered introspection is activated or triggered by the introspective individual’s effort, her genuine interest or her intention, out of either curiosity or concern, in self-probing a current state so as to involve her looking into her stream of consciousness (cf. Ch. IV, §6.2.).

For example, if you are at a concert, many sensations occur within your mind. Introspection selects a specific auditory state as its target in order to learn about the qualitative character of the experience. You are introspecting your auditory sensation of some rustling-like double-bass sound in the opening of the first movement of Shubert’s 8th Symphony (call this auditory state bass), while setting aside or leaving un-introspected another auditory sensation you are having simultaneously of the annoying sound of a plastic bag that your neighbor is noisily crumpling and which, sadly, becomes integrated into the background experience just as the violin enters. Hence, you notice a new auditory

---

94 Ch. II discusses how an individual generates judgments in the mentioned ways.
state with a different content (call this auditory state *bag*) that may be another potential target state of introspection. However, even though both experiences are present, you still select to self-probe your auditory state *bass*. Notice that the object of introspection is the target state at issue with a particular content (auditory state *bass*) rather than the music itself (the sound of double-basses) that is the object of perception.

In emphasizing the selective aspect, some philosophers may contend that introspective awareness does not select a target state—since the target state is simply “caused”—or that the target state does not have any effect on the state in a voluntary or intentionally controlling way but only on what represents the lesser degree of energy for the system. However, I have shown that *stimuli-induced* introspection is just another plausible *mode* of introspective awareness.

Notice that introspectively selecting among mental phenomena implies considering that even if the auditory states *bass* and *bag* are both conscious, the state *bag* can be left un-introspected or set aside as non-probed, thus pointing to or directing us to the target state *bass*—i.e. to the locus of the relevant experience.

Introspection is also selective in the sense that it can begin at the exact point of connecting a conscious aspect of the state *bag* with a non-conscious aspect of that state, and so turn the latter aspect into the target of a subsequent introspective episode. For example, while you are introspectively aware of your auditory sensation of the annoying sound of the plastic bag, that experience might bring to light a feature that suddenly appears, such as a feeling of fury which had otherwise remained non-conscious. So introspection can start by selecting an aspect of the state *bag*, with certain clues upon which the quality of the experience is dependent and, then, can change direction and select another aspect of the same state. Hence, this aspect of the *bag* state then becomes the new target of introspection and the vehicle of your feeling furious.

*(b) Cumulative Process*

Introspection selects from among an array of different mental phenomena in order to probe a mental state. Since introspection learns to recognize the sequence of an introspective episode and
instances of mental phenomena that previously occurred in specific introspective cases—or that match familiar information with new information—it is convenient to characterize it as a *cumulative process*.

Introspection involves a concatenation of mental phenomena as they occur in a series of particular episodes. The introspective process entails a succession: from being in an initial mental state with a certain phenomenology, passing through the generation of an introspective judgment about that state which yields a self-attribution of the state at issue, and ending up perhaps in a report. This succession leads one to obtain a certain familiarity in terms of how introspection proceeds or how the sequence occurs.

But this succession is additionally helpful in recognizing how instances of mental phenomena take place in similar introspective cases: when a mental phenomenon is beginning and when it is coming to a close; when a mental state breaks down (as in an extreme panic attack), interrupts because of a distraction (a loud sneeze), or stops unexpectedly (a pervasive desire); when repetitions of certain sensations or aspects of our phenomenology occur; when an earlier sensation is similar to a current sensation; and so on. In going through the process, introspection gathers certain references to previous occurrences—or, let us say, it collects mental “*leitmotifs*” that persist alongside several introspective cases of a similar sort while organizing them in a sequence.

Assuming that introspection is a selective process *ceteris paribus*, let us now see how it works as a cumulative process. For example, while at home, you are introspectively aware of your auditory sensation of *tremolando* strings in the opening of the third movement of the Sibelius Violin Concerto (call it auditory state *string*) while you notice another auditory sensation is simultaneously appearing. That sensation is engendered by that annoying sound of a plastic bag again, but now it precludes the pulsating string chords and interrupts your introspection of the auditory state string. But this time the sound is caused by your spouse who seems obsessed with plastic bags.

You recognize there is a repeated aspect of your auditory state *bag* that engenders fury when you are introspecting auditory states such as bass or string (call it emotional state *bag*). So a certain

---

95 Recall that *introspective* judgments are ordinarily defined as statements about one’s current mental life, and in self-attributing one judges or makes a judgment about one’s mental life. It is natural to presume that introspective awareness of a visual state of a grasshopper on your taco brings about a judgment on which that particular state with its distinctive phenomenology is based. Specifically, this would be a self-attribution that such a visual experience is happening to you (cf. Ch. IV, § 4.1.1.).
auditory state bag causes an emotional state bag. You recognize the mental phenomena in virtue of previous introspective episodes of a similar sort. You are brave enough to probe the emotional state bag: you are feeling furious again, but now it is more intense. In focusing on the emotional state bag, you detect that your fury is combined with other sensations: it is the crackling rustle of the plastic bag and its revolting smell which annoy you—a turkey sandwich is inside the bag. Also, you detect that your fury intensifies when you think about the indifference of the audience member who opens bags at every concert without considering the existence of the audience, and even more when you think that your spouse forgets your allergy to plastic bags.

This example shows that introspection works as a cumulative process after all. It is by introspection that you learn to recognize the emotional state bag in subsequent introspective episodes. You learn by introspection when an emotional state bag begins, repeats, or finishes, even if every time that you introspect the state you discover similar responses, and sometimes new responses, too.

By considering it as a cumulative process, we can see that introspection can be iterative in probing mental states or aspects of our phenomenology. It accumulates experience or entails experience acquisition—irrespective of the possibility of achieving control of certain mental phenomena in a significant way. Recall that I developed an account of how an acquisition of experience and conceptual resources for the qualities of our conscious experiences may affect the way in which we generate introspective judgments and offer introspective reports. Hence, depending on the amount of practice, constancy and time spent, our introspective ability can become more sophisticated. Introspective judgments of conscious experiences can become more and more refined. I have claimed that we could get introspective competence or expertise in the process itself or in only parts of it (cf. Ch. II, §§2-4).

(c) Predictive Process

Introspection is both a selective and a cumulative process. Since introspection tracks mental phenomena such as episodes of conscious experiences or aspects of its phenomenology, and since it considers previous introspective episodes in anticipating future introspective episodes, it is plausible to characterize it as a predictive process.
Introspection is a predictive process given that it recognizes instances of familiar mental phenomena according to the frequency with which certain types of mental phenomena occur in introspective events of a similar kind. It can make comparisons quickly and can forecast how an experience will continue in a subsequent introspective event while distinguishing it from another. Also, introspection can envision, to a certain degree, what an experience would be like—if, for example, we identified its character, and applied terms for the experience in play; if we distinguished a current experience from a previous experience of similar type in a certain introspective event; or if we possessed the right degree of psychological interest and training.

For example, introspection of your current emotional state bag matches previous similar experiences or episodes of fury, rage, or what you have gone through in previous introspective events of a similar kind; and, in virtue of that, it envisions how you will feel in subsequent introspective events when you encounter the emotional state bag. You can envision that a further emotional state bag can involve a mix of anger and desperation, that it might be accompanied by bodily sensations such as disgust and associations with smell of turkey, or that it might be combined with a belief that elderly people are generally noisy and inconsiderate, or that your spouse is rather fascinated with the sound of plastic bags. Based on specific past events, then, introspection predicts how an experience unfolds in subsequent events of the emotional state bag.

The predictions of subsequent events of the emotional state bag also involve the attribution of extreme annoyance as a property of an experience that has happened to you, and they advance judgments about the emotional states of type bag as experienced by you. If introspection is repeatedly successful in ascertaining the emotional state of type bag in subsequent introspective events, then it is a predictive process. Introspection can thereby also anticipate errors of access by determining whether or not the access will be more accurate. This will depend on the situation, the kind of mental state at issue, or the way in which the individual arrives at self-attributions. If this description of the introspective process is true, introspection can self-correct in either the act of judging states or in the moment of verbally reporting thoughts or experiences.

In maintaining introspection as a predictive process, one might suppose that introspection can change or affect the experience, the content of the experience, the aspects of the phenomenology or
intentionality, the process itself, or one’s further learning, while still engaging in further probing. However, if these impacts were the case, this issue would be a positive influence for introspection, since predictive modes can improve accuracy.

Introspection as a predictive process may affect the way in which we become aware of our subsequent mental states, our improving introspective capacity and our ability to obtain accurate outcomes. Also, it is likely that this would impact our acquiring a familiarity with the introspective process and becoming more introspective individuals.

I have suggested that we can learn to develop introspective capacity with training and practice; this will enable us to predict errors. In other words, with training and practice, an individual will be able to detect and diagnose when she is in error, and where and what the error is. She would thus learn to avoid errors in certain cases and would have the chance to correct them in subsequent situations (cf. Ch. II §3).

Introspection as a predictive process may also be particularly helpful in the detection of sources of error, and in estimation patterns and the prevention of errors. It may help in the generation of hypothesis—beyond what has been accessed—in the evaluation of future responses, the determination of responses to certain problems, and the formation of decisions.

4.3. Variables of Introspection

We currently have no theoretical account that explains in detail how this cognitive process operates or what specific function is at stake. I claim that this new model of introspection postulates conditional cases where certain variables are involved in introspection ceteris paribus, and when both basic conditions of introspection and epistemic conditions of introspective access obtain.

In order to understand how our introspective awareness works, I have claimed that our introspective awareness or introspection is a sui generis process that orchestrates a plural framework that combines and integrates several mental phenomena. In this section, I point out that these mental phenomena can be plausibly identified as variables. I elaborate on what these variables are and why they are called introspective variables.

As I have mentioned, the relevant variables are mental phenomena—any mental state, cognitive process, or resource—that occur while an individual introspectively accesses her mind in a specific time
and in a given situation. It is useful to distinguish between two groups of variables: the first group is suggested by the superposition view, and the second group is taken from the fractionation view.

Among the variables of introspection are: (1) Additional mental states such as bodily sensations that can pop up. These are different from the target mental state or are independent of it, but some of them can be earlier sensed or recently captured while one is probing the target mental state. These additional mental states may be occurring simultaneously or may be a further target state. (2) Mental imagery, which may be visual or auditory, independently sensed, or accompanied by an experience or while one is having an experience that is being introspectively accessed. (3) Background information or information which is derived from previous experiences or from previous introspective episodes. (4) Inner-speech which has been captured as mere words or a set of words that can describe features of the target mental state while it is being introspectively judged. (5) Verbal reports that may or may not accompany introspective judgments. (6) Listening to one’s reports while embarking on further introspection.\(^{96}\)

Among the second group of variables are: (7) captioning (verbal/non-verbal), (8) reintegration, and (9) intensification, all of which have been already been described (§3.3). These are all subject to flexibility and fluctuation along various dimensions, such as scope (possession of conceptual resources or use of conceptualization to label experiences in an introspective episode), level (accuracy in encoding information) and degree (intensity or richness of detected experience). Let us call these sub-variables. It is worth providing some examples of how sub-variables can figure in introspection. For instance:

Scope: The individual psychologically captions experiences in an introspective episode from which she is smelling something and conceptualizes “this smell is good,” after which she passes through the judgment “I am smelling the tomato,” and then perhaps generates a self-report, such as I adore this smelling sensation.\(^{97}\)

Level: The individual psychologically reintegrates information coming from her sensory system: her olfaction system connects the current mental state with a particular content, such as the smelling of a tomato, and from a certain level of accuracy encodes this mental state for storage so that it can be recalled in working memory.

\(^{96}\) Space prevents me from discussing work in progress on listening to one’s verbal reports and the difference that this can make in introspection. This will be the subject of a future study.

\(^{97}\) Prinz (2004) uses this example in a different way.
Degree: The individual psychologically intensifies the degree of vividness, richness or detail detected in a particular conscious state, such as the sensation of smelling the tomato, after which her attention is increased and her sensory system is rendered available to working memory.

Now, these mental phenomena (1-9) are called variables because they are prone to variation based on, at the very least:

(A) The specific situation and spatial-temporal framework: a certain context, following a certain succession, and the particular duration of an introspective episode.

(B) The way in which they appear or show up: whether alone with the target state or in tandem and in specific combination with other variables. E.g. the result of an introspective episode may be different when there are verbal reports as opposed to when the individual remains silent.

(C) The target mental state at issue: introspection generates different judgments depending on the kind of mental state at issue, such as whether it is cognitive or non-cognitive.

(D) What the content is of the mental state bringing about the differences in the introspective episode. E.g., when the content changes from a desire to p to one to q (see §4.4.1.).

(E) The difference across individuals in terms of self-ascriptions at different moments—i.e., how specifically an individual self-ascribes a particular mental state, resulting in a particular introspective episode. Recall that for introspection what is important is how the individual experiences a particular experience rather than what the experience is actually like.

Some clarifications are important. Variables can causally interact and render different results. Also, variables happen while the introspective process is ongoing, even if there are some mental phenomena that remain non-conscious. In scrolling over different variables occurring simultaneously, one can “suspend judgment” (roughly, à la Husserl) about many of them, or about the significance of a particular current experience, in order to gain introspective access to the conscious experience itself. Alternatively, one can self-trigger only one experience while inhibiting another occurring experience.

Notice that while it might be the case that all variables occur in introspection, often only some of them occur. In fact, it is not necessary that all variables occur in introspection, but only that some of them do. To be clear: it is not that through variables we are aware about the target mental state, but that some or all of those variables occur when we introspectively access our mental states and engage in the
introspective process. It may be the case that introspection can occur when only one variable is instantiated. In this case, the variable would interact with the target mental state at issue.

Variables also can reveal the cooperation of diverse information occurring in the introspective process. These comprise properties that affect the introspective process in a relevant way. That is, different mental features may be revealed at different stages of the information processing or in certain properties of mental states. For instance, conscious experiences can serve as useful data for generating other mental phenomena which are not part of the initial target state, such as bodily sensations.

Consider the conductor analogy with our introspective awareness. During a symphony there may be an unaccompanied violin solo. However, the conductor is still guiding the concert, and can introduce the other instruments and coordinate them with the strings. So, even when only one variable is present, it is guided by a process that is disposed to integrate it with the rest.

To illustrate the core of this section, see the following diagram:

**Mental phenomena** that are integrated within introspection

**Targets**
- Mental states (cognitive/non-cognitive states)

**Upshots**
- Self-attributions or judgments (see Ch. II for grains of introspective judgments)

Targets and upshots of introspection are constant mental entities referred to epistemic conditions of access

**Variables of introspection**
1. Additional mental states occurring simultaneously
   a. or popping-up while introspecting
   b. e.g. bodily sensations
2. Mental imagery
3. Background information
4. Inner-speech while judging
5. Verbal reports
6. Listening to one’s verbal reports while reporting and embarking in further introspective episodes

**Variables and sub-variables**
7. Captioning (verbal/non-verbal)
8. Reintegration
9. Intensification

Scope of possession/use
Level of accuracy
Degree of richness

Inputs to the process
Outputs to the process

164
4.4. The Function of Introspection

I claim that mental phenomena are combined and integrated in superposition—i.e., a particular relation of variables or interaction among of different variables—in virtue of an additive effect and a proportional result, I shall turn to this below. The additive effect and proportional result characterize the specific kind of integration that is at stake.

Recall that introspection is a way to be aware of our own current and recently past mental states and to self-attribute those mental states. Introspection focuses on concurrent states as they occur in our mind or in the stream of consciousness, and enables us to describe and report our thoughts and experiences (cf. Ch. I).

So far, the proposal is that in being aware of a target mental state, introspection operates in tandem with other cognitive processes. But its specific function is to combine and integrate different mental phenomena (variables), which is possible on the basis of several phenomena occurring. Thus, introspection is not reducible to any one of these phenomena understood separately, but is sui generis.

In order to explain how several mental phenomena can be combined and integrated, it is useful to appeal to the notion of superposition as it is found in certain sciences. This notion can also serve to show how introspection possesses the advantage of preserving distinctions among these mental phenomena.

‘Superposition’ explains the behavior of several systems; it is used in mathematics (for linear systems), physics (for wave systems), quantum mechanics (in describing wave behavior and registering the propagation of certain type of waves), and geology and archeology (for the strata states identified and the layers set down in a time sequence).

Let us take only a couple of these characteristic explanations, in their simplest form, to see how application of the principle of superposition to both linear systems and wave behavior can make the idea clear.
A linear system is a simple mathematical model that supposes that a cause-and-effect relation is always proportional. For example, when we apply a force of 1 by stretching a spring and it is deformed to 10, we say that the cause (stretching it) is proportional to the effect (the deformation of the spring) by 1:10. If we stretch it out, we obtain a constant relation between the force and the deformation. Given a linear system, the sum of causes is proportional to the sum of effects.

**Stretching a spring**

\[
\begin{array}{ccc}
Force & Deformation \\
\rightarrow 1 & \rightarrow 10 \\
\rightarrow 2 & \rightarrow 20 \\
\rightarrow 3 & \rightarrow 30
\end{array}
\]

the sum is identified as superposition

Figure 3 is intended to illustrate the principle of superposition, which refers to the sum of forces (causes) and deformation (effects).

Waves are typically described as variations of amplitude through space and time—the height in a water wave, the pressure in a sound wave, or the electromagnetic field in a light wave. In any system with waves, the waveform at a given time is a function of the sources—e.g., external forces create or affect the wave. In many cases, the equation describing the wave is linear, and thus the superposition principle can be applied. That is, the net amplitude generated by two or more waves traversing the same space is the sum of the amplitudes that would have been caused by the individual waves separately—e.g., two waves traveling towards and then passing right through each other without there being any distortion on the other side.\(^98\)

Although my use of ‘superposition’ as applied to introspection is somehow independent of these specific applications, there are some resemblances. The theory of superposition has at least two options to offer when it comes to successfully explaining the mental phenomena that occur in introspection.

Superposition is a function—i.e., a relation of variables—constituted by both (1) an *additive effect*: the sum of variables yields the superposition of responses, and (2) a *proportional result*: a change in variables supposes a change in the result. For instance:

\(^{98}\) ‘Superposition’ also refers to the case of Schrödinger’s cat: an experiment in which a cat is alive or dead depending on the emission of a particle, but it is indeterminate until one looks the result; I bypass discussion on this point.
(1) An introspective episode can carry either the sum of all variables (mental phenomena) or only some of them; and we can say that the variables that figure in it are superposed. These mental phenomena are probed as they occur or appear to the introspective mind. Introspection can organize them irrespective of which one, in fact, comes first, or which triggers another—e.g., whether a conscious experience is prior to a belief or whether the belief indeed entails images or sentences.

(2) An introspective episode can carry a specific result according to the particular variables that it entails. A change in variables (mental phenomena) supposes a change in the upshot or deliverance of introspection; likewise, a richer input will produce a richer output. In fact, when there is an increase in variable intensification, the sub-variable degree of richness will result in turn in an increase in the accuracy of the introspective judgment (cf. Ch. II.).

When focusing on mental states that manifest an abundant stock of contents, the relation among certain variables can render the process longer, but perhaps deeper, bringing about diverse findings: nuances in phenomenological features, nuances in the distinctions between cognitive and non-cognitive states, and, eventually, prompts which can encourage further inquiry to seek the underlying causes of certain mental states.

To understand such a function (superposition), it is important to take into account that the combination or integration of variables translates into a dynamic process. Its development as an additive ability and the results it provides enable us to learn about the mind, and most importantly, to learn about the conditions under which introspection operates.

It is plausible to claim that introspection works as a superposition of several variables that are combined and integrated during a certain time and a given situation. The resulting interaction of variables can be called superposition. Since ‘superposition’ is applied to the particular mental phenomena that occur in introspection, it is convenient to label this function cognitive superposition. In other words, given the interaction of variables, it is sound to maintain that mental phenomena such as mental states, cognitive processes, and resources are combined and integrated in superposition. Alternatively, we might say that the mental phenomena will be cognitively superposed in introspection.

To illustrate what superposition can do, consider the following example showing how introspection can provide information about a specific target mental state. I will posit a desire that turns
into an intention when other processes such as perception and inference cooperate (see also “change” in the next section). Introspection can provide information about a target mental state such as an emotion and, after acts of attention and memory take place in the introspective event, the target state can change to a belief. In the next section, I will offer a simpler example in which several mental phenomena occur and are superposed in a particular introspective episode.

Before showing how this notion about how superposition would work in a particular introspective episode, it is worthwhile to inquire first why it would be necessary to appeal to ‘superposition’ when several contemporary theories allude to a certain integration of information. That is, the integration of information from many sources is an important feature of consciousness. Content integration in particular plays an important role in various theories, especially in global workspace theory.

Recent theories of consciousness studies tend to focus on mental mechanisms that are unifying. For example:

(a) Cognitive broadcasting theories focus on when consciousness is generated by a global access to selected bits of information in the brain and nervous system (e.g., Global Workspace Theory) (Baars, 1988, 1997; Dehaene et al., 2000).

(b) Information integration theories focus on when consciousness is determined by integrated information; here it is possible to get information on the parts of a system and its organization (e.g., Integrated Information Theory) (Tononi, 2008; Tononi and Edelman, 1998).

(c) Temporal binding theories focus on when one’s consciously perceiving an object implies our visually seeing its multiple parts and features but nevertheless experiencing the object as a unity (e.g., the Binding Problem) (Singer, 1999; Crick and Koch, 1990).

(d) Cognitive unison theories focus on when consciousness is treated as a phenomenon within which information processing gets integrated and organized (e.g., The Cognitive Unison Theory of Attention) (Mole, 2011).

Although the contribution of these theories may be significant for either the development of a philosophy of introspection or for their eventual application to introspection, these accounts are specific to consciousness and primarily allude to working memory and attention. None tackles the specific question of how introspection of our mental states arises.
Moreover, although these theories may have some points in common with my own theory—for instance, (a) in the suggestions that cognition integrates information from several sources; (b) that there is a mechanism of integration of information; (c) that there is a mechanism for binding information; and (d) that breaking cognition up into its parts or simpler mechanisms helps to see how the parts work together—there are fundamental differences between this group of theories and my theory.

As for the notion of integration, according to my theory it is the function of superposition that integrates several mental phenomena in introspection—that is, whereas in (a) and (b) it is working memory which plays the role of integration, and in (c) and (d) it is attention which plays the role of integration. Additionally, my notion of integration is not intended to always entail information or contents, as in (b), but is meant to indicate, precisely, an integration of the processes that cooperate with introspection. My notion of integration does not involve binding items such as words that seem to have separable letters, as in (c), nor does it always bind features or properties. It rather instead involves an integration of the mental phenomena that occur together.

Additionally, in my theory, the items at issue are in superposition, whereas in (a) the items at issue are in competition for attention. Moreover, my theory is not that introspection integrates mental states in general, but rather that it integrates certain kind of mental phenomena (see diagram §4.3).

While other theories tend to say, in different ways, that consciousness integrates things in general, according to my theory there is an integrator of a specific kind: one that superposes diverse mental phenomena so that we can be aware of our mental life and arrive at self-attributions. Recall that it superposes in virtue of both an additive effect: that the sum of variables yields the superposition of responses; and a proportional result: that a change in variables supposes a change in the result.

What is it about the integration that introspection does that is distinctively introspection-like? It superposes several mental phenomena as they occur in an introspective episode while enabling us to determine the character of a particular ongoing mental state that is the vehicle of a particular content. Recall that it might be the case that all variables occur in introspection—though often it’s only some of them that occur.
4.4.1. An Introspective Episode: Cognitive Superposition at Work

I aim here to show how a hypothetical introspective individual can be aware that she is in a mental state and can determine what kind of particular ongoing mental state it is, including its contents—i.e. which particular conscious experience or attitude is the vehicle of a particular content \( p \) or \( q \)—when several mental phenomena in superposition are occurring simultaneously in a specific introspective episode. Additionally, I contend that a practiced introspective individual can identify changes, transitions, and boundaries between mental states in specific introspective cases.

I have suggested that introspection works as a superposition of several variables by which we come to understand important aspects of our mind. Since different mental phenomena are superposed in the introspective process, the introspective individual can organize her introspective judgments about the target mental state even when other mental phenomena occur simultaneously.

Although the description of the following introspective episode uses the form of a sequence, it is important to bear in mind that the hypothetical scenario below is not intended to suggest that introspection is a sequence of mental phenomena. It is rather intended to illustrate how cognitive superposition accounts for the way in which an individual self-probes her mental states and arrives at self-attributions.

Consider this case: [1] you have been tossing and turning for hours. Though you have tried everything, you cannot sleep. You turn to the last alternative: counting sheep. You continue counting for quite some time, and find doing so painfully boring. You suspect that this experience is common, but that it usually works nonetheless (background information). [2] Finally, you are becoming sleepy...you have a sensation of relaxation as you begin to doze off (introspective target mental state). [3] This sensation is accompanied by the representation of a facial expression of yourself smiling (mental imagery). [4] You say to yourself: “Yes, I did it!” (inner-speech). That naïve inner utterance, expressing a slight “claim of victory” or “wrong inference,” is enough to trigger an inquiry into your ongoing experience within your stream of consciousness. (Unfortunately, you had to remember: “Don’t count your chickens before they hatch.” Not only does your inner speech make you lose track while counting sheep, but your last-ditch remedy for insomnia has failed, and you are condemned to stay up all night).
[5] You feel anxious (*bodily sensation*) while you mentally visualize dark circles under your eyes (*mental imagery*). [6] You ask yourself, in inner speech: “What the hell is happening to me?” (*captioning*). [7] This inquiry leads you to understand something else about your mind. And, perhaps with further probing, you also end up learning more about some of the possible causes of your insomnia. For example, you were fed up with feeling strangled by your turtle-neck sweater, or you were resisting Silenus’s warning that “a life spent in ignorance of one’s own woes is most free from grief…”

[8] Insofar as you have as an object of introspection a target mental state—your sensation of relaxation—you are able to attribute to yourself this state. You may judge that *your sensation of being on the verge of sleeping is exquisite* or that *your tiredness is the result of having not gone to bed earlier*, and that accumulating many thoughts does not help with sleep. Insofar as during the introspective event you recognize some other mental phenomena (variables), it is likely that you are able to organize those mental phenomena and articulate your judgments in a relatively clear way. Introspective judgments consider some mental phenomena that have taken place in superposition.

Suppose that at this moment you are able to verbally report your mental state to another person. You are able to report your experience while being introspectively aware of the differences of your state with their respective phenomenal character.

Accordingly, [9] you may express aloud a thought like: “I suddenly felt so relaxed and calm, but something distracted me again, and now I am upset and so tired! I feel completely desperate and helpless. I believe that I cannot deal with these feelings any longer, and I am afraid that they are starting to bleed over into the rest…”

[10] Verbal reports here are considered as another variable of introspection. The verbal reports screen the information gained throughout the introspective episode and crystalize how the individual organizes her introspective judgments about the mental phenomena or occurrences in a relatively systematic way.99

99 If an experimenter were to form a question about the state at issue, it might be a precise question for which a much more limited response would be expected. On the difference between concurrent and retrospective reports, and reports of an actual experience unbiased by interpretation, see Piccinini (2004). I bypass discussion on this issue. Subjective reports of conscious experience have also played an important role in the search for the “neural correlates of consciousness” (Rees & Frith, 2007; Tononi & Koch, 2008; Prinz, 2012; see also Varela 1996).
Perhaps we should look at first-person reports to explain the cognitive superposition. For example, your experience could be a mix of desperation and anger combined with another mental state, such as a belief about your failure. Here, you are able to report your mental states while being aware of their differences with respect to other mental phenomena—e.g., certain bodily sensations like feeling anxious or mental imagery, such as visualizing dark circles under your eyes.\(^{100}\)

To illustrate the core of this scenario, see the following diagram:

**Mental phenomena occurring in an introspective episode:**

- Target mental state [2]
- Background information [1]
- Mental imagery [3]
- Inner speech [4]
- Captioning) [5]
- Verbal Reports [10]

\[\begin{array}{c}
\text{Additive} \\
\text{Effect}
\end{array}\]

\[\begin{array}{c}
\text{Proportional} \\
\text{Result}
\end{array}\]

\[\begin{array}{c}
\text{Introspective Judgments}
\end{array}\]

\[\begin{array}{c}
\text{Superposition of mental phenomena}
\end{array}\]

To analyze how introspection works, we need to take into account the combination and integration of the introspective target mental state. Along with other mental phenomena occurring in specific introspective events, I have called this kind of process *cognitive superposition*.

The superposition theory maintains that for any target mental state that is introspectively accessed, it is possible to determine its character and content—i.e., what the target mental state is, and which conscious experience or attitude is the vehicle of its particular contents—and also to distinguish the additional mental phenomena, such as visual imagery and inner speech, that figure in a specific introspective episode.

Notice that the episode just described indeed exists in time: it is temporally extended and has certain duration, even though we do not register this with exactitude. That is, when one is trying to sleep, all those mental phenomena occur in time—at the most a few moments before or after the introspection of the target mental state. Also, mental phenomena seem to occur within the same temporal framework as events in the environment, even though their exact timing is not easy to determine or describe.

\(^{100}\) For similarities and distinctions between perceptual content and mental imagery content, see Nanay 2015.

172
The previous scenario represents the progression of a stream of consciousness in an introspective episode when one is simultaneously aware of a feeling of desperation about not being able to sleep. Also, it shows that in order for an individual to be aware of that particular mental state, it is necessary that she be conscious with respect to a relative time sequence with a particular duration: meaning, specifically, that she is having the experience of insomnia.

Features [1-10] are representations of the mental phenomena that can occur during an introspective episode while an individual is introspectively accessing the target state as the phenomena are being activated or prompted by introspection.  

Now let us see how cognitive superposition can account for certain properties of the mental states at issue. That is, if we have diverse mental phenomena in superposition, and if the introspective individual can detect its character and content, she can also be introspectively aware of her states as having certain properties. This view maintains that after learning, training, and practicing, it is possible to identify changes, transitions, and boundaries between mental states in accordance with the specific time and situation.

Introspective awareness can reveal, at the very least:

- The change or shift of direction from a recently past mental state to a current mental state within a specific introspective event (ms\(^1\) at \(t_i\), ms\(^2\) at \(t_2\)), or the change in qualitative character of the same experience within a specific introspective event (\(e_q\) at \(t_1\), \(e_q\) at \(t_2\)).
- The transition from an occurring mental state to a fading or dissipating mental state as a specific introspective event (from ms\(^1\) to ms\(^0\) at \(t_i\)), or the pathway from having a powerful experience to including it within an introspective event (from \(e^1\) to \(e^0\) at \(t_i\)).
- The boundary between two joint mental states occurring at the same time within a specific introspective event (ms\(^1\)–ms\(^1\)’ at \(t_i\)), or between two similar mental states having different qualitative character within a specific introspective event (ms\(^1\)–ms\(^2\) both of \(q\)).

\(\text{\textsuperscript{101}}\) In a future study, I will examine how new states that are triggered can count as part of the same introspective episode while engendering further introspection.
Now, let us see how a practiced introspective individual can identify changes, transitions, and boundaries among mental states in specific introspective events. We will see how my model sheds light on the phenomenological analysis presented here.

For simplicity’s sake, let us focus on the introspective awareness only, and leave out how an individual explicitly generates introspective judgments about such states and their properties. She can introspectively distinguish mental states and their properties as varying, passing through, and delimitating, while they are appearing to the mind.

- While you are probing your visual perception of a bottle of mezcal and go in depth into the qualitative character of this experience, you notice a sudden change in your target mental state: your visual perception has just shifted into the desire to have a glass (or two) of mezcal. Alternatively, while you are probing your gustatory sensation of the mezcal and enjoying its strong smoky flavor, you find out by introspection that the character of your experience suddenly changes when you see a worm in it.
- While you are probing your auditory state of hearing yourself laugh and enjoying the bodily sensation that it brings about, you detect by introspection the transition in the state from a chuckle, to a smile, to a feeling of embarrassment; and all this precisely at the time you notice you are laughing with a horrible screeching sound.
- While you are probing your auditory sensation, which you savor, of simultaneously crying and laughing, you detect by introspection the boundary between this sensation and the emotional pleasure that accompanies it. Either that, or, while you are probing your sadness, you discover that the boundaries between this sadness expressed by crying are, on the one hand, sorrowful, or, on the other hand, desperate.

**The superposition theory accounts for:**

Introspective awareness can reveal, at the very least

After learning, training, and practicing, introspection can also reveal

- Character
- Content

of the target mental states

- Changes
- Transitions
- Boundaries

between mental states
What is relevant here is that the introspective individual is aware of being in certain mental states as they change sequentially, transition between themselves, or maintain the boundaries between them.

Introspection is a flexible process that can respond to the intensity of one’s experience and to the changes of its quality, or in identifying one’s changes of attitude and one’s accompanying mental states. Thus introspection centrally involves perception or attention. Moreover, introspection can respond to the changes of a visual state by appealing to accurate memories. Thus it can be extremely responsive to fluctuating stimuli by its ability to switch from short-term memory to retrospection. The same flexibility applies in identifying the transitions and boundaries between the other mental processes that accompany introspection.

5. Conclusion

After discussing the single- and plural-process models of introspection, I have proposed a new model that provides a plausible explanation of the operation of introspection.

I have shown that, although a specific view of the plural-process model is more compelling—namely, the fractionation view—it does not analyze specifically how the parts of introspection are connected. It is also the case that proponents of other models of introspection have not discussed what specific relation between introspection and several mental phenomena is involved.

I have built on the specific plural-process view that lends itself to examining how certain cognitive processes and resources can accompany introspection. I have argued that Introspection is not simply a representation of a current mental state: rather it works in tandem with different cognitive processes and resources and, most importantly, it combines and integrates different mental phenomena (variables).

I have suggested the term “superposition” as a useful notion to explain the specific relationship or function that introspection may entail with its variables (I have called it “cognitive superposition”). I have argued that this function can underpin or fit within an anti-reductionist position. I have explained how mental phenomena work in specific introspective cases. I have suggested that “cognitive superposition” generates a new and precise account of how introspection works or the way in which we become aware of our mental states and arrive at self-attributions. Although some of the details of the superposition
theory remain to be worked out in future research, as far as I know, no one as yet has shown the applicability of “superposition” for explaining the operation of introspection.

I have shown that introspection is a cognitive process that operates in superposition with several mental phenomena by which we become aware not only of the character and contents of our mental states, but also of the changes, transitions, and boundaries between them—in particular, after learning more about their character, receiving training about what to derive from them, and then practicing introspection because of them. The outcome of this proposal should be not only the expansion of the plural-process model, but also a supplement to those alternative philosophical and scientific theories of introspection that are currently in development.
Bibliography


Rey, G. (2012). *Postscript* to “We Aren't All Self-Blind: A Defense of a Modest Introspection” https://sites.google.com/site/georgesrey/


