SUMMARY

Do I believe that it is raining? Do I believe that my partner and I will grow old together? Do I intend to pay back the money I borrowed? Do I prefer strawberries over raspberries? Do I value family over work? Should I focus on having fun, being a parent, a career woman, a good friend? These kinds of questions, both the more trivial and the more substantial ones, are central to this dissertation. They are the kinds of questions whose answers, if true, provide one with a piece of self-knowledge, namely, self-knowledge of one's own intentional mental attitudes.

Self-knowledge of mental attitudes is often regarded as special because such attitudes involve a specific first-person perspective, namely a commitment to what the attitude is about. For instance, if I believe that it is raining then I take it to be true that it is raining. And if I value family over work, I take family to have more importance than work. This connection between attitude and commitment, and especially the kind of self-knowledge that “respects” this connection, is the focus of this dissertation. I call this kind of self-knowledge transparent self-knowledge. It is transparent because it is a mode of knowing one’s mental attitudes in which one, in a way to be specified, looks beyond or through the attitude to what the attitude is about. The underlying question addressed in this dissertation, which consists of five independent papers, is: how should transparent self-knowledge be conceptualized?

My study of transparent self-knowledge has been partly born out of amazement at the increasingly all-encompassing scientific perspective on the nature of the human mind. From the scientific perspective, all self-knowledge is regarded with skepticism and sometimes even declared illusory. To give an idea of this position, consider the following characterization of our capacity for self-knowledge by Daniel Dennett:

...each of us is in most regards a sort of inveterate auto-psychologist, effortlessly inventing intentional interpretations of our own actions in an inseparable mix of confabulation, retrospective self-justification, and (on occasion, no doubt) good theorizing. (Dennett 1987, 91; emphasis in original)
Self-attributions of mental attitudes, in this view, are nothing more than the result of theorizing about what could go on in our minds that would explain what we do. Theorizing that is often better left to persons other than ourselves, for our own theorizing is obfuscated by our self-conception.

The departure point of this dissertation is to question the scientific skepticism about self-knowledge and its underlying assumptions about the nature of self-knowledge. What seems to be most problematic about the view of self-knowledge purported by science is the relation between a person and her own mental life that it presupposes. Their skepticism about self-knowledge addresses the idea of a person who must take on the role of a psychologist and who then tries to observe what is going on in her own mind. Like a bystander who merely witnesses what goes on in her head. As opposed to being a witness or bystander, a person doesn’t merely register what is present in her mind. Rather, her mental goings-on express her view of things, i.e., what she takes to be true, what she will do, how she feels about things and what she wants. Her mental attitudes should thus be seen by her ‘as expressive of [her] various and evolving relations to [her] environment, and not as a mere succession of representations (to which, for some reason, [she] is the only witness’) (Moran 2001, 32). Seeing one’s attitudes as expressive of one’s stance on the world at large implies that a person doesn’t relate to her mental life as a psychologist but as someone who is actually inhabiting the perspective purported by those attitudes – as someone who is committed to the various things inherent in holding different mental attitudes. Given that transparent self-knowledge respects this connection between attitude and commitment, it can be seen as the alternative to the alienated scientific view of self-knowledge that has led to so much doubt and suspicion regarding the human capacity to know our own minds.

What is transparent self-knowledge?

Back to the underlying central question of the dissertation: how should transparent self-knowledge be conceived? First of all, let me relate it to philosophy of self-knowledge more generally. A principal point of departure in thinking about self-knowledge is the difference between knowledge of one’s own mental attitudes and the mental attitudes of others. This difference is predominantly viewed in light of its epistemology: self-knowledge is thought to be privileged, i.e. to have a more secure epistemic status, and to be available through peculiar access, i.e. through means available only in knowing one’s own mental attitudes (cf. Byrne 2005; Gertler 2015). What is often left out of these discussions of the epistemology of self-knowledge is the connection between self-knowledge and the nature of the person who is seeking
self-knowledge. Why should self-knowledge matter to us? What are the connections of self-knowledge to personhood, to moral psychology, and to (mental) agency? In what way do these moral psychological issues inform the difference between self-knowledge and knowledge of someone else’s mental attitudes?

A philosopher who has refocused attention in the philosophical debate on self-knowledge to such moral psychological issues is Richard Moran (2001). The essential difference between knowledge of one’s own mental attitudes and the mental attitudes of others is, according to Moran, not a difference in privileged or peculiar access but a difference in the way a person is involved in her own mental life. Different from the relation a person may have to someone else’s mental life, her relation to her own mental attitudes is the aforementioned first-personal agental stance. She isn’t some expert witness, but inhabits the perspective purported by her mental attitudes. This is to say that in believing that it rains she takes it to be true that it is raining; in feeling hurt she commits herself to the view that someone wronged her or something is hurtful; in intending to go to the new Wes Anderson movie tonight, she commits herself to making it her business to be there.

In respecting the relation between attitude and commitment, transparent self-knowledge seeks to do justice to this essential difference between self-knowledge and knowledge of someone else’s mental attitudes. It is self-knowledge that, in the case of belief, not merely puts me in a position to report that I have a certain belief, but ‘to speak of [my] conviction of the facts’ (Moran 2001, 76). Similarly, in the case of intention, in being aware of my intention to go to the movie tonight, I am not aware of some likelihood that I will end up at the movie theater tonight. Rather, I am aware of having it made my business to be there. Hence, the idea is that transparent self-knowledge is the kind of self-knowledge that a person has from the stance of agency. That is, as someone who has an active relation to her own mental attitudes: both to whom it matters what her mental attitudes are and who is involved in what attitudes she holds.

The limits of transparency procedures: its scope
Having given a first characterization of transparent self-knowledge is not yet to give an account of transparent self-knowledge. This would minimally require addressing the question how such transparent self-knowledge is to be achieved. It is generally thought that achieving transparent self-knowledge should be explicated as going through a transparency procedure. Broadly speaking, such a procedure can be called transparent if it involves answering an inward-directed question about one’s mental attitude by answering the relevant outward-directed question about the content of the attitude. For instance, one can achieve self-knowledge of one’s belief that $p$, if
one answers the question whether one believes that $p$ by answering the question whether $p$. One main objective of this dissertation has been to analyze such transparency procedures. What I have argued is that these procedures face two kinds of limits: limits in scope and in solving what I have dubbed the Two Topics Problem (TTP). Let me discuss both in turn, before coming back to the relation between a transparency procedure and transparent self-knowledge.

Are there limits in the scope of transparency procedures? In the dissertation, I have looked at the scope of Moran’s transparency claim. Chapter 1 begins by addressing the question what Moran’s transparency claim precisely consists of. Before being able to evaluate the claim that a person is to answer the question whether she has a belief that $p$ by answering the question whether $p$ is true, we need to know what is required in order to answer that latter question. Moran’s work supports three different requirements: 1) that there aren’t any conditions on how to answer the question whether $p$; 2) that one should refer to reasons in favor of $p$; and 3) that one should refer to reasons justifying $p$. These three requirements are evaluated by checking whether each of them holds for numerous examples of belief, ranging from recalcitrant beliefs to beliefs based on no evidence. Take, for instance, Elisabeth’s belief based on non-justifying reasons. Years after the war is over, Elisabeth keeps on believing that her husband will return home from said war. When asked why she believes this, she cites as her reason that there have been other men who have returned home after the war, although she is also aware of many countervailing reasons. Since her belief isn’t based on justifying reasons, she cannot meet the third requirement: she cannot answer whether it is true that her husband will return home by reference to reasons justifying that proposition. However, the fact that her belief isn’t based on justifying reasons doesn’t preclude her having any reason in favor of the proposition believed. Hence, she might fulfill the second requirement. As such, she can also meet the first requirement, because that requirement attaches no conditions to how the question about the proposition is to be answered.

Based on expositions and considerations such as these, Moran’s transparency claim seems most plausible, i.e. has the widest scope, if it means that the first requirement should be met. This isn’t very surprising, of course, because the first requirement is fairly minimal. But because it is so minimal, it also has some counterintuitive results. It actually seems to be in tension with a person’s active relation to her own mental life – to what Moran calls the deliberative stance. The exposition of the case of Elizabeth shows that she can meet the first requirement without taking a deliberative stance: she might stubbornly repeat the proposition believed without really being open to the question whether to believe that
proposition. This tension between the first requirement and the deliberative stance is also evidenced in another class of belief, namely obsessive beliefs. Even if a person only has reasons against believing that she will fail the exam, she might, when faced with the question ‘Will I fail the exam?’, cannot but think that she will. This means that she might meet the first requirement, although her relation to this belief bears no signs of the active relation that Moran seeks to incorporate in his account of self-knowledge. The upshot is that friends of Moran’s transparency account either need to give up the deliberative stance (which was supposed to be its main motive) or justify its limited scope with respect to distinct classes of belief.

Another question about the scope of transparency procedures is whether they apply to mental attitudes other than belief, such as emotions, desires, cares, etcetera. Chapter 4 takes up the question how Moran’s account could be translated to mental attitudes other than belief. It assumes that Moran’s transparency account works for belief and then seeks to apply it to emotion. The basic difficulty in such application is that where the relevant “outward-directed” question for belief is simply whether the proposition under consideration is true, the relevant “outward-directed” question for emotion is less easy to discern. The reason for this is that emotions do not only seem to be about the world, but also about what is important to the person having the emotion. Even if we all agree that a person has betrayed me, I need not feel betrayed if either the person or the betrayal itself were insignificant to me. Similarly, only if I care about a sports team, will their wins and losses spark joy or disappointment, respectively. We only feel an emotion if something matters to us (cf. Helm 2010). Chapter 4 thus argues that Moran’s transparency claim cannot be applied to emotions, at least not without incorporating an account of the relation between transparency and what matters to us (a question that is addressed in Chapter 5).

Hence, Moran’s delineation of a transparency procedure faces several limitations in its scope. First, it either fails to apply to several classes of belief or it becomes minimalistic to such a degree that it seems to be disconnected to a person’s active relation to her mental attitudes. And secondly, it doesn’t apply to emotions, nor, presumably, to other mental attitudes that are similarly connected to what matters to us.

The limits of transparency procedures: the two topics problem
This brings us to the second limitation of transparency procedures, namely the limit in solving TTP. The basic idea of transparency procedures is that a person answers a question about her own mental attitude by answering a relevant question about the topic of that mental attitude. The question is, however, why answering the latter
question is related to answering the former. What puts me in a position to answer the question whether I *believe* that it is raining by answering the questions whether it *is* raining? There seems to be something outright puzzling going on here, because the fact that it is raining doesn’t entail that I believe that it is raining, nor provides evidence for it: one can imagine numerous scenarios in which it is raining but I do not believe that it is raining or in which I believe that it is raining but it isn’t. Hence, transparency procedures face TTP: the problem that the truth of $p$ doesn’t seem to provide an epistemic basis for the truth of $I\ believe\ that\ p$.

A careful glance at the state of the debate on transparent self-knowledge shows that there is no consensus of what the relation between $p$ and $I\ believe\ that\ p$ might be, nor what kind of solution respects the commitments of transparency views that actually establish the source of TTP. Chapter 2 seeks to provide a grasp on the nature of the different responses to TTP. The responses that I discuss are: 1) the view that TTP is only apparent; 2) inferential views; 3) judgment views; and 4) metaphysical views. In very general terms, the proceeding arguments are as follows.

First, I argue that TTP has to be accepted as a genuine problem insofar as one accepts the transparency intuitions that in self-ascribing a belief that $p$ a person both makes an empirical claim that she is in a certain state of mind and endorses $p$. Secondly, taking Alex Byrne’s (2018) account as exemplary for the inferentialist response, I contend that a crucial assumption in his account, namely that inference from a premise entails belief in that premise, is unwarranted (a claim that is corroborated in Chapter 3). Thirdly, I argue that both the judgment views and the metaphysical views need, albeit for different reasons, to presuppose a form of attitudinal awareness, i.e. an awareness of one’s judgment or belief regarding $p$. This is incompatible with delineating a transparency procedure to achieve self-knowledge. Hence, transparency procedures find their limit in TTP.

However, I don’t think that this means that achieving transparent self-knowledge is impossible. We need to distinguish clearly between transparent self-knowledge and a transparency procedure to arrive at self-knowledge. Where the former expresses something about the nature of the self-knowledge at issue, the latter aims to identify an epistemically justified method of moving from thinking about the world at large to a self-ascription of a mental attitude. My explication of the nature of transparent self-knowledge as self-knowledge that manifests the connection between mental attitude and view of the world, doesn’t imply that it necessarily depends on a transparency procedure. Even if the way in which one achieves transparent self-knowledge is not fully transparent, there is no immediate implication that there isn’t another way in which such self-knowledge might be achieved. This is where I would like to point to the prospects of the involvement of
some form of attitudinal awareness, which was inherent in many of the accounts discussed in Chapter 2.

Since transparent self-knowledge is a mode of knowing in which one inhabits the perspective purported by the attitudes, invoking a form of attitudinal awareness to explain how we arrive at such self-knowledge brings with it a certain tension. After all, being aware of the attitudinal aspects of one’s mental attitude implies that one isn’t focused solely on the content of those attitudes. How could transparent self-knowledge be compatible with a form of attitudinal awareness? Such compatibility would seem to depend on the way in which this form of awareness modifies a person’s relation to her own mental life. It shouldn’t impugn on, most importantly, the relation between attitude and commitment. Put differently, we might say that the compatibility of transparent self-knowledge and the involvement of attitudinal awareness requires the latter to meet certain transparency requirements. For instance, the requirement that any uncertainty in one’s self-ascription should direct one’s attention back to the content of one’s self-ascribed state (i.e. gazing back at the world). In my view, future research on this topic should thus focus on which transparency requirements attitudinal awareness should meet if it is to be compatible with transparent self-knowledge.

The form of self-knowledge

If transparency procedures face all these difficulties, one might wonder why again we needed transparent self-knowledge in the first place. The basic motivation for thinking there must be such a thing as transparent self-knowledge is to be able to make a genuine distinction between the first-person and third-person perspective (and thereby between self-knowledge and other-knowledge). What sets the first-person stance apart is, first and foremost, the connection between having a mental attitude and grasping the world in a certain way. What is especially important about transparent self-knowledge is thus that a person sees her mental attitudes not as part of the passing show but as expressive of her commitments. Seeing her attitudes as expressive of her own stance implies that the question of taking responsibility has become pertinent to all of them. When self-ascribing an attitude, we either take or fail to take responsibility for our attitudes by avowing or disavowing them. Avowing an attitude consists of a self-attribute of the attitude including an explicit endorsement of its content. In the case of belief, to avow my belief that \( p \) is to express my commitment to \( p \)'s truth. Because of this, achieving the kind of transparent self-knowledge at issue involves manifesting one’s agency: for only if a person takes the responsibility to avow her mental attitude will she achieve transparent self-knowledge.
Safeguarding this stance of agency in matters of the human mind is thus, as argued in this dissertation, the reason why we need transparent self-knowledge. Nonetheless, different problems remain to be addressed. One worry is that a person’s active relation to her mental attitudes is more present in the case of trivial attitudes (such as one’s belief that it is raining) than in the case of substantial mental attitudes. Substantial mental attitudes, such as one’s cares, concerns, deep desires and values, seem to be too integrated in one’s entire life to depend merely on one’s taking the responsibility to avow them. Were a person, say Katherine, to desire another child, this should be reflected not only in her avowal on the matter but also in a wide range of actions and reactions (cf. Lawlor 2009). It thus seems that transparent self-knowledge loses its relevance in the case of substantial attitudes.

In Chapter 5 I take up this challenge and argue to the contrary: even if such patterns of action and reaction form part of coming to know my substantial mental attitudes, avowing these attitudes remains essential and has a unique status in coming to know them. My arguments show that the status of avowal is unique, first of all, because the significance of patterns of action and reaction, and what such patterns tell about our attitudes, ultimately depends on avowal. Secondly, they show that avowal is essential to knowing one’s substantial mental attitudes, because these attitudes require one to have a self-conception. Acquiring self-knowledge of substantial mental attitudes can be seen as a struggle to fulfill the commitments pertaining to these attitudes. A struggle that requires a person to manifest her agency – to take responsibility for who she is and putting herself at risk of being challenged and making mistakes. In this view, avowal is, par excellence, essential to substantial self-knowledge, not only to trivial self-knowledge.

A final worry for providing a genuine conception of transparent self-knowledge is the wide variety to which it is supposed to apply. Following the literature, it seems as if we should slightly adjust the conception of transparent self-knowledge for each different mental attitude. How are we supposed to account for some kind of unified conception of transparent self-knowledge? The suggestion I want to make is that the analytic Aristotelian approach would be helpful to answer this question. I have introduced analytic Aristotelianism in Chapter 3, in a discussion on the nature of reasoning. Reasoning, like self-knowledge, comes in a wide variety. Chapter 3 develops an argument against the claim that all reasoning necessarily involves a change in attitudes (this argument also undermines the assumption that is required by Byrne’s transparency procedure, as discussed in Chapter 2). Although it seems obvious that reasoning often involves such a change in attitudes, e.g., forming, revising or withdrawing a belief, that doesn’t imply that a change in attitudes is necessarily involved in reasoning. For instance, we quite often
reason hypothetically or merely check the validity of an argument, without having determined for ourselves whether we believe the premises. As Wright (2014, 28) has put it, we should ‘ distinguish inference in general from coming to a conclusion…; no particular attitude to [a] proposition is implicit in inference itself.’ By discussing examples of reasoning without a change in view, it will become clear that a different approach to reasoning is needed: namely, one that includes instances of reasoning with and without change in attitudes.

The alternative view that I develop is the form view. It holds that all the instances of reasoning can be unified if we adopt the view that reasoning is first and foremost one thing: namely recognizing the truth-connection between two statements (i.e. making a judgment that \( p \) as following from \( q \)). This is the kind of thing that reasoning is; its logical form. The logical form of a concept consists of the form of thought or the form of judgment that underlies the concept and it refers to what can be predicated of the thing in question, say \( X \). The logical form (or structure) is revealed by analyzing the things that can be said or asked about \( X \), and thus by analyzing our practices and abilities regarding \( X \). I have argued that we need to understand reasoning in this way especially because of two reasons: first, it seems impossible to analyze reasoning in terms of sufficient and necessary conditions (or in terms of an explanatory essential feature or property), and secondly, because this seems to be the only way in which the extensive variety of instances of reasoning form a unity.

The suggestion I want to make is that something similar is true of transparent self-knowledge. I think that understanding the nature of transparent self-knowledge – understanding what it means to have self-knowledge of, say, one’s belief that \( p \) in which one takes \( p \) to be true – requires an analytic Aristotelian approach. This accords with Moran’s remark that

\[
\text{...for a range of central cases, whatever knowledge of oneself may be, it is a very different thing from the knowledge of others, categorically different in kind and manner, different in consequences, and with its own distinguishing and constraining possibilities for success and failure. (Moran 2001, xxxi)}\]

Future research of transparent self-knowledge should therefore, in my view, elaborate on the idea that the distinctiveness of transparent self-knowledge lies in its logical form. Such an inquiry would require an analytic Aristotelian approach.

*Self-knowledge, science, and agency*
I conclude this summary by returning to the opposition between a third-personal observational approach of self-knowledge that thrives in the scientific domain and a first-personal agential approach. The former approach rests on the assumption that mental attitudes can be inferred from a person’s patterns of action and reaction (Chapter 5). But is it really possible to ignore the subject’s active relation to her own mental life and the agency manifested in taking and having a stance? What I have argued is that, even if some mental attitudes might be thus inferred, each instance of inferring depends on the subject taking a stance. Mental data, inner promptings, and patterns of action and reaction have symptomatic value only if, somewhere down the line, the subject takes them to be expressive of her perspective. Mental data and internal promptings aren’t just given facts about the subject that she can discover, experience or note, but are themselves an expression of the subject’s commitments to the world at large. They are, at least always partly, an expression of her agency. It thus seems that we cannot do without the first-personal agential perspective. This gives us reason to think that a wholly scientific third-personal perspective of the mind cannot be adequate.

By saying that science cannot bypass the agential perspective, I don’t mean to claim that my approach couldn’t benefit from closer attention to scientific results. This would be a worthwhile and exciting line of future research. The point I want to make is that a genuine understanding of the nature of our mental life requires us to reject a wholly mechanistic view of our minds. As Philippa Foot puts it in her Moral Arguments (1958, 509):

...evidence is not a sort of medicine which is taken in the hope that it will work... When given good evidence, it is one’s business to act on it, not to hang around waiting for the right state of mind...

Foot’s remarks are spot on. Evidence, reasons, inner promptings, and mental attitudes require the subject to take a stance – to manifest her agency. Rather than waiting for the right state of mind, the reflections presented in this dissertation are an invitation to take a stance; an invitation to reconsider the importance of transparent self-knowledge and the indispensability of a thoroughgoing philosophical analysis of self-knowledge.