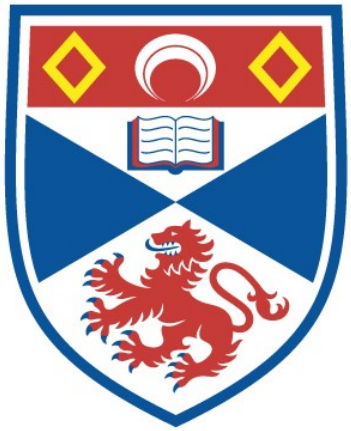
I'm a believer : Evans' transparency remark and self-knowledge

Paul John Conlan

A thesis submitted for the degree of PhD
at the
University of St Andrews



2022

Declaration

Candidate's declaration

I, Paul John Conlan, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 61,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in February 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date Signature of candidate

7/2/22 (Paul Conlan)

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

7/2/22 (Adrian Haddock)

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as

requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Paul John Conlan, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date Signature of candidate

7/2/22 (Paul Conlan)

7/2/22 (Adrian Haddock)

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Paul John Conlan, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date Signature of candidate

7/2/22 (Paul Conlan)

Table of Contents

Abstract

The central goal of this thesis is to understand self-knowledge through understanding a particularly difficult and promising remark of Gareth Evans', from his *The Varieties of Reference* (Evans, 1982), a remark which has formed the basis of so called 'Transparency' accounts of self-knowledge. Evans' Transparency Remark is sometimes read as deflationary of self-knowledge in some respect, and I hope to show that although Evans' account is indeed deflationary of our ordinary idea of self-knowledge, it retains what we might consider central features of an account of self-knowledge.

I do this by giving an overview of the literature surrounding Evans' remark and making a distinction between Rationalist and Inferentialist accounts of Transparency. I also suggest that the goal of an account of self-knowledge is to explain, or explain away, the phenomenon of Privileged Access. Having done this we return to Evans' development of his remark, and from that I develop a novel Rationalist account of self-knowledge of belief which hews closely to Evans' own development but differs in one significant way, which leads to an answer to one of the central objections to Transparency accounts of self-knowledge, the Puzzle of Transparency. Having developed this Simple Account of Transparency and defended it against what I take to be the major objections to Transparency accounts, I turn to the best developed Inferentialist account, Byrne's *Transparency and Self-Knowledge* (Byrne, 2018), and suggest why we might find his account wanting. Finally, I suggest ways in which the Simple Account of Transparency might be extended into a general account of self-knowledge, and suggest there is one important unanswered question remaining.

Acknowledgements

General Acknowledgements

There are a number of people without whom this thesis would not have been possible.

Academically I am deeply indebted to Dr Adrian Haddock, Prof. Peter Sullivan, and Prof. Crispin Wright, my supervisors for this project. Adrian's course focussed on Sellars' *Empiricism and the Philosophy of Mind*, which I took as a third-year undergraduate, fundamentally shifted my view on philosophy. He has since supervised both my undergraduate and Mlitt dissertation as well as this PhD thesis. He has taught me that the hardest thing in philosophy is saying what you mean to say and has challenged me to properly formulate my thoughts at every stage of my work. Peter provided invaluable supervision during the dissertation's difficult middle period, and without his careful directing I would never have appreciated Strawson and Evans' subtlety and brilliance. Both Adrian and Peter's supervision looms large over the final thesis here, and without them I would never have completed this work.

I am also indebted to the other members of the Stirling philosophy department and to the Knowledge Beyond Natural Science project members, in particular Prof. Crispin Wright, who has been instrumental in my coming to a clear view regarding Alex Byrne's work, and to a clear view of the idea of 'privileged access' discussed in the thesis. Dr Giovanni Merlo and Dr Giacomo Melis, the postdoctoral researchers on the KBNS project also deserve mention here for their support of the graduate students on the project. Prof. Jim Pryor and Prof. Alex Byrne also commented on early versions of chapters contained herein in their capacity as project auditors, and I am grateful for their constructive and helpful criticism.

I have presented parts of this thesis at various work-in-progress seminars, and I have been asked piercing questions by interlocutors too numerous to mention. To them I say thank you, this work is stronger for your challenges.

I am also indebted to Indrek Lobus, Jose Mestre and Lachlan Devine for various philosophical discussions which have deeply impacted the final version of this thesis and, I hope, further work. Finally, I wish to thank my thesis examiners, Dr Lucy Campbell, and Prof. Michael Wheeler – their thorough and constructive engagement with this thesis and the suggested additions and clarifications have made this a stronger piece of work.

There are of course several people who contributed to my finishing of this work who did not provide academic support. Foremost amongst these are my partners. Roslyn and Elisa, without your love and

support I would never have made it through the writing process. You have both been the stars by which I steer my course, and without you, I am nothing.

My mother, Jan, sister Janine, stepdad Ali and nieces Caila, Ailsa and Eilidh have provided invaluable support throughout my PhD; there is no way I could have completed this without you all.

Numerous friends have helped in various ways; Jon, Lou, the Graeme's, Jared, Lee, Mhairi, Gabriel, JJ, the RPGnet folks: Thank you all. Writing this thing has been a marathon, and you all have been cheering me on the whole way. Stuart Hay deserves special mention as the only one of my friends outside academia who cares about the ''I' stuff.' There definitely aren't two things known, Stu.

Finally, this thesis is dedicated to the memory of my grandmother, Arlene Wallace, who sadly passed away before she could see me complete it. Without her, I would never have been able to pursue academic philosophy. Thank you, Gran. This is for you.

Funding

This work was supported by the Templeton Foundation *Knowledge Beyond Natural Sciences* project [grant number 58450]

Introductory Remarks.

Precis of the Thesis

The central goal of this thesis is to understand self-knowledge. Specifically, its goal is to understand self-knowledge through understanding a particularly difficult and promising remark of Gareth Evans', from his *The Varieties of Reference* (Evans, 1982), a work otherwise not focussed on epistemology, but on philosophy of language. This remark has motivated what are commonly called 'Transparency' accounts of self-knowledge and is sometimes read as deflationary of self-knowledge in some respect. I hope to show by the end of this thesis that although Evans' account is indeed deflationary of our ordinary idea of self-knowledge, it retains what we might consider central features of an account of self-knowledge – Privileged Access understood as Authority and Groundlessness. I do suggest in the account developed and defended in chapters three and four that Evans' account is deeply radical and perhaps revisionary of our capacity to know our own beliefs, and that a proper appreciation of Evans' remark will show how self-knowledge can be both substantial and no cognitive achievement.

Chapter one introduces the main puzzling question regarding self-knowledge and develops ways in which that question might be framed. This will act as a guide to the greater discussion within the thesis. In particular, I clarify the idea of Privileged Access which is often taken to be the central phenomenon in need of explanation in an account of self-knowledge. The main business of the chapter, however, is to give an overview of ways in which Evans' remark has been understood in the literature, dividing the developments of Transparency accounts of self-knowledge into a Rationalist strand and an Inferentialist strand. Finally, the four standard objections any account drawing on the Transparency Remark must face are detailed.

In chapter two, I return to Evans, giving a detailed discussion of his account of self-reference, and the Strawsonian (and Kantian) background which leads to the Transparency Remark. In this chapter we also see the insight which forms the basis of the account developed in chapter three emerge from the discussion of Evans' remark.

The third chapter presents a new, Rationalist, account of the Transparency Remark, inspired by Matthew Boyle's work in *Transparent Self-Knowledge* (Boyle, 2011), but hewing much closer to Evans' original insights. This is (what I call) the Simple Account of Transparency, which forms the central original contribution of the thesis. In presenting this account, I give a relatively austere explication of Evans' remark, which is in central ways truer to Evans' own ideas than Boyle's development. I do, however, depart from Evans in one vitally important way which makes the Simple Account more than merely an exegesis of Evans' remark. There is a standard objection to

Transparency accounts of self-knowledge — what I call the Puzzle of Transparency — and a significant amount of the business of this chapter is spent on clarifying what this objection amounts to, answering the objection and engaging with some consequences of the answer provided. I show how a proper understanding of the role of the first-person in assertions leads both to a dissolution of the Puzzle of Transparency, and to a radically revised notion of what Evans' remark entails. This radical development of the Transparency Remark draws heavily on Anscombe's (1981) discussion in *The First Person* and it is in this way that the Simple Account diverges sharply from Evans' own account. I also show how the account can fend off the central objections detailed in chapter one, with the exception of the Objection from Scope, which is discussed in chapter four.

Chapter four centrally deals with the Objection from Scope by discussing Matthew Boyle's *Two Kinds of Self-Knowledge* (Boyle, 2009). In this chapter I aim to understand Boyle's idea that a certain sort of self-knowledge is 'fundamental'. Boyle's contention is that, if the self-knowledge he defends as fundamental is indeed just that, then a uniform account of self-knowledge is not in the offing. My aim is to agree that the self-knowledge Boyle is interested in is indeed fundamental (and broadly in the way he suggests) but to argue that this does not rule out a uniform account of self-knowledge. Rather, I suggest that we can still have a uniform account, by shifting our focus to the *form* of the account of self-knowledge Evans' remark suggests, rather than its content. By the end of this chapter, I will have defended the Simple Account against the four central objections to a Transparency Account – The Objection from Scope, the Objection from Over-Intellectualisation, the Objection from Anti-Luminosity, and the Puzzle of Transparency. I will also have suggested how the Simple Account might be the basis of a general account of self-knowledge.

The fifth chapter engages with the central thesis of the most well developed unified Inferentialist account of self-knowledge which purports to be a development of Evans' Transparency Remark — namely, that defended by Alex Byrne in his *Transparency and Self-Knowledge* (Byrne, 2018). Byrne presents the most promising Inferentialist alternative to the Rationalist Simple Account. The main goal of this chapter is to understand Byrne's proposal, and to pinpoint some areas where Byrne might have to do further philosophical work, and then to present one substantial objection to the account. This objection presents Byrne with a trilemma which suggests that Byrne's account can either be an Inferentialist account, or a development of Evans Transparency Remark or an account which is neither inferential nor a development of Evans' Transparency Remark, but is rather a causal-reliabilist account of self-knowledge.

Finally, in chapter six, I make some concluding remarks on the achievements of the thesis, flag one important unanswered question, and suggest a promising direction for further work.

1. A Taxonomy of Transparency

In spite of their brevity, Gareth Evans' (1982) remarks in *The Varieties of Reference* regarding how one knows what one believes have inspired a significant strand of approaches in how to make sense of self-knowledge.[1]. As Evans puts it:

> "If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting in to operation whatever procedure I have for answering the question whether *p*." (Evans, 1982, p. 225)

Accounts which are inspired by or aim to develop Evans' remarks are understood as *Transparency* accounts, as Evans' remark suggests that the question of what one believes is *transparent*[2] to the question of what is the case. As such, we will refer to the above remark from Evans as the Transparency Remark. Unpacking just what the Transparency Remark means and how it might inform a conception of self-knowledge is the central business of a Transparency account, and in addition to Evans' own remarks, there is a substantial literature of attempts to make good on Evans' central idea. The aim of this chapter is to give an overview of the literature whose goal is to develop or understand the Transparency Remark and give a brief account of the central objections Transparency accounts must face. As such, the chapter will proceed very schematically in giving an overview of the literature that is concerned with accounts inspired by Evans' remark. Before proceeding to the literature overview, however, we should consider what the aim of an account of self-knowledge developed from Evans' remark is.

1.1. The Puzzle of Self-Knowledge

Self-knowledge is (*prima facie* at least) different from our knowledge of worldly matters, and there are a number of ways of elucidating these differences. An account of self-knowledge should give us either an explanation of why self-knowledge is different from knowledge of the world — what I shall call 'other-knowledge' — or an account of why the apparent difference between self-knowledge and other-knowledge is no difference at all. But to do this, we must have a grasp of what the intuitive idea that our knowledge of ourselves is different from our knowledge of the world is actually picking up on.

[1] Evans is not the first to suggest such a view, Hampshire (1975) suggests something similar, as does Wittgenstein in the Blue and Brown books, and Moran credits Edgely (1969) with helping to inspire such a view. Nevertheless, Evans is the origin of the modern idea of a Transparency account of self-knowledge.

[2] The major point of divergence between Transparency accounts is, as we shall see, in the understanding of what such 'transparency' of one thing to another amounts to.

1.1.1. Privileged Access

Self-knowledge *prima facie* exhibits *Privileged Access*. For example, our access to our own beliefs and desires has — in a sense that needs to be explained — 'better credentials' than our access to the beliefs and desires of others. To take an example drawn from Byrne (2018, pp. 5-8): Jim is in a better position to tell whether he wants a cup of tea, or whether he believes it is raining than he is regarding whether his officemate Pam desires that or believes this. As Byrne puts it:

> "…Jim's claim that he wants tea would usually be treated as pretty much unimpeachable, whereas his claim that Pam wants coffee is obviously fallible. (Jim's evidence points in that direction: Pam normally has coffee at this time, and is heading to the office kitchen. However, she drank her coffee earlier, and now wants a chocolate biscuit.)" (Byrne, 2018, p. 7)

Jim's access to his own desires in some way differs with respect to fallibility than his access to Pam's, and part of the question of self-knowledge is what this difference amounts to.

Giving a sharp characterisation of Privileged Access is somewhat difficult, but we can pull two notions from Byrne's short example above: Jim's claim about his own state has better epistemic credentials than his claim about Pam's state, and the basis or grounds for Jim's claim about his own state are distinct from those for his claim regarding Pam's[3]. We can partition the notion of Privileged Access into two phenomena that when taken together are constitutive of Privileged Access: *Authority* and *Groundlessness*[4]

1.1.1.1. Authority

We can see, in the example of Jim and Pam above, the idea of Authority. As Byrne puts it, Jim's claim is treated as 'unimpeachable'. That is, when a speaker makes a self-ascription, their ascription has better epistemic credentials than the ascription an interlocutor would make. When a speaker makes a self-ascription, the speaker is on the face of it correct in what they aver if the ascription is a genuine one[5].

---

[3] It would be easy to suggest here that the intuitive idea of grounds here doesn't get to the point; of course, the grounds are different: one claim is grounded in how things are with Jim, the other in how things are with Pam. This, I think, misses the point in question. It is not merely the specific grounds or evidence Jim appeals to in making his claim, it is the type of grounds or evidence *in general* that Jim appeals to. This will be discussed in more detail below.

[4] Wright (2015) suggests a third 'cornerstone' of Privileged Access, *Salience*.

> "…selves tend to know what there is to know—selves' mental attributes do not, in general, elude their awareness" (p.50)

I think it is unclear, given a sharp characterisation of Groundlessness/Immediacy what work we can put Salience to that is not done by Groundlessness. As such, I will not discuss Salience further.

[5] An example of a non-genuine ascription in these terms would be an obviously hyperbolic or sarcastic one – the speaker can be clearly understood to be engaging in non-genuine speech (distinct from lying, for example).

1.1.1.2. Groundlessness[6]

We can also see in the example above that Jim comes to his claim about whether Pam wants a coffee in a different way from the way he comes to his claim that he wants a tea. To come to his claim that Pam wants a coffee, Jim looks to some available evidence (Pam is on her way to the kitchen) and infers that Pam wants a coffee. Jim does not seem to have to look to evidence in the same way to come to the claim that he wants a tea. Rather, Jim's state is in some sense to be specified directly available to him; he does not need to observe himself to know his own state; indeed, he need not base his self-attribution on any evidence whatsoever[7]. Coliva (2016, pp. 52-58) suggests that, given the sorts of concerns we canvassed regarding Jim and Pam, an account of self-knowledge should understand self-knowledge as neither based on observation nor based on inference, and such an account would then be an account of self-knowledge as groundless. Coliva does, however, go on to distinguish the idea that self-knowledge is groundless, in the sense just outlined, from the idea that self-knowledge has no basis.

> "A word of caution is in order. Groundlessness consists in the idea that self-knowledge is not the result of any substantial cognitive achievement, such as observing or inferring from a symptom to its likely cause. It does not necessarily involve the idea that one's psychological self-ascriptions are not based on anything, such as the very experience one is undergoing when, for instance, one is in pain and avows it." (Coliva, 2016, p. 54)

So, the suggestion that groundlessness is no cognitive achievement is not the suggestion that it is what Cassam would call 'insubstantial'. When considering why self-knowledge is not inferentially grounded, Cassam writes:

> "One possibility is that inference can be neither required nor relevant when it comes to self-knowledge because knowledge of one's own thoughts is normally based on nothing. But this is hard to swallow. Maybe cognitively insubstantial self-knowledge—say the knowledge that I am here—can be based on nothing, but self-knowledge isn't cognitively insubstantial and so can't be based on nothing." (Cassam, 2014, p. 123)

Given the characterisation of groundlessness as the claim that self-knowledge is neither observationally nor inferentially grounded, an Inferentialist account of Transparency is committed to the denial of groundlessness, and instead of explaining privileged access in terms of groundlessness,

---

[6] Groundlessness is sometimes called *Immediacy*; I will use these terms interchangeably.

[7] Consider the concerns Jim might point to in answer to the questions 'Why do you believe Pam wants a cup of coffee?' and 'Why do you believe you want a cup of tea?'. In the former, Jim will point to the evidence available regarding Pam's behaviour. In the latter, Jim is just as likely to answer 'because I want a cup of tea'. Cases like this can give the impression that for self-attributed states, one need only point to the state itself as the grounds for one's attribution (which would be a form of groundlessness). I suggest that while there is something right about this, we should treat such examples with a little suspicion; as we go forward, we will see that Evans' remark gives a way of understanding these cases such that they are not as such groundless, but where the idea that Jim's grounds for his belief about himself and his grounds for his belief about Pam are still distinct in a substantial and interesting way.

the Inferentialist must instead explain away the purported groundlessness of self-knowledge. A key reason for the Inferentialist to resist groundlessness is to resist the idea that self-knowledge is insubstantial, in Cassam's sense. As Cassam forms the criticism, the insubstantiality of self-knowledge is a matter of its being based on *nothing*, and self-knowledge cannot be based on nothing. What I have suggested above, however, is a notion of groundlessness that does not amount to the idea that self-knowledge is based on *nothing*. Perhaps we can hold on to a notion of groundlessness while also holding on to the idea that self-knowledge is not based on nothing[8]. Throughout this section I will remain neutral on whether self-knowledge is groundless or groundlessness must be explained away, but in chapter three I will present an account of Evans' remark which is groundless in Coliva's terms.

1.2. A General Account of Self-Knowledge

Along with an explanation of Privileged Access (or an explanation of why we might jettison Privileged Access) we should consider the *scope* of an account of self-knowledge. Is our account of self-knowledge *general* (or *unified*)? That is, does it explain all self-knowledge (or provide a systematic formula for doing so) under one rubric? Alternatively, is the account a divided account which explains only part of self-knowledge, such as knowledge of belief, or knowledge of sensation, without explaining self-knowledge in any other domain. The Objection from Scope (discussed in section 1.4.4. of this chapter) is an objection levelled against accounts which are not general, suggesting that such accounts are insufficient to explain self-knowledge satisfactorily. As I will be taking the Objection from Scope seriously, we will consider the provision of a general account of self-knowledge to be an explanatory goal to be achieved.

1.3. Transparency Accounts

We should begin by discussing Evans' own development of the Transparency Remark. We begin by giving a brief contextualization of the remarks as part of his larger project, given their place in a text that is fundamentally about *reference*, not about the epistemology of self-knowledge. Having done so, we will further carve up the territory of those discussions that follow Evans into a Rationalist strand, where the move from the answer to the question 'is it the case that *p*?' to the answer to the question 'do you believe that *p*?' is a matter of a rational reflection of some kind and an Inferentialist strand[9], where the answer to the world-directed question is a premiss or basis for an

[8] We might think of the apparent groundlessness of self-knowledge in terms of answering 'why-questions' regarding our beliefs, and what we can point to in answering the question 'why do you believe that'. An account of self-knowledge that was 'insubstantial' in Cassam's sense would either refuse this question or would not allow the subject to point to anything relevant in answering this question.

[9] I take this distinction to be inspired by Cassam (2014)

inference to the question of belief.[10] While I suggest it would be correct to place Evans in the Rationalist camp of Transparency accounts, for now we will hold him apart as we discuss the genesis of such accounts.

1.3.1. Evans

The Transparency Remark suggests a (*prima facie*) deflationary account of self-knowledge in which the epistemic procedure for giving the answer to the question 'do you believe that *p*?' is the same as the epistemic procedure for giving the answer to the question 'is it the case that *p*?' (Evans' remark is put in terms of being 'in a position to answer a question', but this is a subtlety that can at this stage be ignored). Evans does not fully develop this epistemological strand of his insight beyond suggesting the following rule:

> "We can encapsulate this procedure for answering questions about what one believes in the following simple rule: whenever you are in a position to assert that p, you are *ipso facto* in a position to assert 'I believe that *p*'." (Evans, 1982, pp. 225-226)

Central to Evans' understanding of this procedure and of self-ascription in general is the idea of the Generality Constraint – the constraint that sentences (and *inter alia* thoughts) observe a certain structure:

> "If we hold that the subject's understanding of 'Fa' and his understanding of 'Gb' are structured, we are committed to the view that the subject will also be able to understand the sentences 'Fb' and 'Ga'." And we are committed, in addition, to holding that there is a common explanation for the subject's understanding of Fa' and `Ga', and a common explanation for his understanding of 'Fa' and 'Fb'." (Evans, 1982, p. 101)

The upshot of this is the idea that a subject who can assert 'I believe that *p*' must possess (as Evans puts it) "…the psychological concept expressed by 'δ believes that p' , which the subject must conceive as capable of being instantiated otherwise than by himself" (Evans, 1982, p. 226) where 'δ' is a 'generic' self-concept (or 'I-idea'). Evans' discussion makes significant use of this notion of an 'I-idea' and the role of self-ascription in an account of self-knowledge in an attempt to fully respect the Generality Constraint in his discussion. A detailed exposition of Evans' account will be given in chapter two of this work.

[10] The distinction between 'rationalist' and 'inferentialist' is intended to exhaust the possibilities for the shape of an account that makes good on Evans' remark, with the caveat that the possibility space for a rationalist account is considerably larger than for an inferentialist account – rationalist accounts take a variety of shapes, with their focus being on a certain sort of capacity on the part of the self-knower, whereas inferentialist accounts focus (broadly) on the mechanism that allows for the movement from the world-directed question to the question of belief.

1.3.2. Inferentialist Accounts

An Inferentialist account of Transparency tries to make sense of the idea that the question 'do you believe that *p*' is (in some sense) answered by answering the question 'is it the case that *p*?' through the idea of some kind of inference.

Gallois (1996) aims to give an account of self-knowledge with a focus on Authority motivated by the Transparency Remark. He suggests that the move from the answer to the world-directed question, 'is it the case that *p*?' to the answer to the question of belief, 'do you believe that *p*?' is via a special form of inference, captured by what Gallois calls *The Doxastic Schema*:

p

So I believe that p

This inferential schema is motivated by considering Moore's Paradox, which suggests there is an absurdity in asserting a statement of the form '*it is raining* but I don't believe that *it is raining*.' Gallois uses this absurdity to suggest the following: "…there is something amiss with my saying, or thinking, that p, but also saying, or thinking that I do not believe p." (Gallois, 1996, p. 46). From the recognition that something is amiss with this assertion, Gallois derives the Doxastic Schema. Of course, the inference described by the Doxastic Schema is a manifestly bad inference, nothing about *p* being the case entails that one does, or even that it is likely that one does, believe that *p*. As such, the fact that it is raining does not suggest one should infer that one believes it is raining. After all, one might never have looked out of their window to consider the weather.

Byrne (2018)[11] takes Gallois' Doxastic Schema and uses the idea of an inference of this form to develop a fully-fledged account of self-knowledge which promises to account for a subject's knowledge of her mental states in general, for example, their beliefs, desires, intentions, sensations and perceptual states. Byrne presents the best worked out Inferentialist account of the Transparency Remark. Byrne's account attempts to substantiate how the inference described by the Doxastic Schema from Gallois could be a good one, suggesting that making sense of this inference is the puzzle that the prima facie obvious insight of the Transparency Remark presents us with. Byrne's suggestion is that the proper understanding of the structure of the inference is in terms of following an *epistemic rule*, which is a conditional of the following form:

(R) "If conditions C obtain, believe that *p*" (Byrne, 2018, p. 101)

[11] Byrne (2018) provides a substantial extension of earlier work contained in (Byrne, 2005) and (Byrne, 2011).

The canonical example of an epistemic rule which produces self-knowledge would be the belief rule BEL:

> "If *p*, believe that you believe that *p*" (Byrne, 2018, p. 102)

Following an epistemic rule such as BEL is captured by the following schema below (substituting the consequent of the particular rule for the variable *p* in the schema):

> "…S follows the rule R ('If conditions C obtain, believe that *p*') on a particular occasion iff on that occasion:
>
> (i) S Believes that *p* because she recognises conditions C obtain
>
> which implies
>
> (ii) S recognizes (hence knows) that conditions C obtain
> (iii) Conditions C obtain
> (iv) S believes that *p*" (Byrne, 2018, pp. 101-102)

Substituting BEL for the generic epistemic rule in the schema gives the following:

> S follows the rule BEL ('If *p*, believe that you believe that *p*') on a particular occasion iff on that occasion:
>
> (i) S Believes that she believes that *p* because she recognises that *p*
>
> which implies
>
> (ii) S recognizes (hence knows) that *p*
> (iii) *p* obtains
> (iv) S believes that she believes that *p*[12]

Condition (i) in the schema above suggests that S forms the belief in question *because* she recognizes C-conditions obtain, and the 'because' here is understood to mark a "…reason-giving causal connection…" (Byrne, 2018, p. 101)

Byrne contends that BEL is a good rule of inference because it is *strongly self-verifying*; trying to follow BEL will produce (by and large) true beliefs, even in the cases where the antecedent condition of the rule is not met — because it is not the case that *p*, and as such the subject does not recognize but merely believes that *p* — and as a result,, the outputs of BEL are *safe* beliefs (insofar as they could not easily have been wrong). Byrne considers the beliefs formed as a result of following (or trying to follow) BEL being safe beliefs a sufficient condition for those beliefs to amount to knowledge. As the premier Inferentialist account, I engage in substantial detail with Byrne's account of belief in chapter five of this work. As his account of belief is the central, or canonical example of

[12] In substituting BEL into the generic rubric for an epistemic rule, we have not replaced the variable *p* with a concrete example. Nevertheless, we can easily substitute our concrete example *it is raining* and see how S, upon following BEL reaches the belief that she believes that *it is raining*. We will continue to substitute the variable for our concrete example in what follows.

his Inferentialist account of Transparency, his account stands or falls with that discussion, without considering his larger project to bring all self-knowledge under the epistemic rubric of epistemic rules, and as such the discussion offered in chapter five is narrowly focussed on the case of belief and the framework of epistemic rule following in general[13].

Cassam (2014) offers an alternative Inferentialist account, based on the idea that an account of self-knowledge should be an account for *us*, fallible, human knowers, not what Cassam calls 'homo-philosophicus':

> "Just as behavioural economists distinguish between *homo sapiens* and *homo economicus*, the idealized rational agent of so much economic theorising, so I think it is helpful to distinguish real humans as we know them from *homo philosophicus*, the idealized subject of so much philosophical theorising." (Cassam, 2014, p. ix)

Cassam calls the difference between *homo sapiens* and *homo philosophicus* 'the disparity' and giving an account of self-knowledge which respects the disparity between the ideal rational agent and the fallible human agent is the central aim of Cassam's *Self-Knowledge for Humans*. Cassam gives a simple statement of his Inferentialist approach as follows:

> "[S]uppose you know that you have a certain attitude A and the question arises how you know that you have A. In the most straightforward case you know that you have A insofar as you have access to evidence that you have A, and you infer from your evidence that you have A. As long as your evidence is good enough and your inference is sound you thereby come to know that you have A." (Cassam, 2014, p. 138)

Note that this suggests that self-knowledge is not groundless or immediate insofar as Cassam suggests self-knowledge is based on evidence. Cassam is explicit about his denial of the idea that self-knowledge is groundless. Cassam suggests self-knowledge is either: (1) Based on inference, (2) Based on Inner Sense or (3) Based on nothing. He suggests that he gives reasons to reject Inner Sense[14], and that self-knowledge is based on nothing should be rejected lest self-knowledge be "…cognitively insubstantial, like knowing that I am here now". (Cassam, 2014, p. 140) Self-knowledge, he suggests, is not insubstantial like this. So, if (2) and (3) are ruled out, Cassam suggests we need to accept (1), that self-knowledge is based on inference, on pain of not knowing our own minds. Cassam suggests that, because (1-3) exhaust the alternatives in giving accounts of self-knowledge, so as long as we can give an inferential account that is in good standing, then it should be taken as the model for self-knowledge.

[13] For more general critical discussion of Byrne's project, see (Conlan, Merlo and Wright, 2020)
[14] We will not go over these reasons here – we are interested in Cassam's positive view, not his arguments against alternative views of self-knowledge.

Cassam takes this to have ruled out self-knowledge being groundless or immediate. However, given how Cassam constrains inference, a *sense* of groundlessness or immediacy on the part of the subject may still be available. Cassam suggests that the inference by which one comes to self-knowledge need not be conscious. The inference can be understood as an unconscious psychological or cognitive process which moves between contents. Further, Cassam suggests that this notion of a movement between contents might not even be the correct characterisation of 'inference', rather 'inference' is a reference to the justificatory structure: "…your knowledge that P is inferential if your justification for believing P comes in part from your having justification to believe other, supporting, propositions." (Cassam, 2014, p. 139)

Regardless, the point is that a sense of immediacy or groundlessness could perhaps be recovered on Cassam's Inferentialist account, if we understand the groundlessness of self-knowledge to be in terms of evidence the subject is aware she possesses. A subject *S*'s belief that *p* could be groundless insofar as there is no evidence *S* is aware of that stands in support of *p*, even if such evidence exists. So, the inference Cassam thinks underpins self-knowledge might be groundless for *S*, and as such groundless in some sense, even if evidence is available.

Cassam argues by example for the inferential view, suggesting that Inferentialism gives a tidy explanation of *S*'s knowledge of her desires in cases where her desires are not obvious (she infers from her behaviour that she does in fact desire something) and in cases where she knows 'straight off' what she desires. Cassam suggests this is the difference between conscious and unconscious inference. The second strand of Cassam's argument for inferentialism draws on experimental evidence from social psychology, which point to the conclusion that

> "…in every case, you have a subject who, in keeping with the postulates of SPT [Self-Perception Theory], knows his own opinion or attitude by inference from his own behaviour. The subject doesn't just know." (Cassam, 2014, p. 147) [Clarification mine]

While Cassam agrees that the experiments discussed are not enough to secure conclusive proof of inferentialism, he does suggest that they weaken the argument against the idea (which Cassam attributes to Ryle) that subjects infer their internal states from their own behaviour. What Cassam suggests is rather that what these experiments show is that the Rylean view isn't quite as dead a duck as it is made out to be, that "…while the evidence for self-attributions needn't be behavioural, it certainly can be." (Cassam, 2014, p. 148) Cassam also suggests that inferentialism neatly explains Moran's idea of the 'presentational view', where "…many statements of the form 'I believe that P' are […] nothing more than 'the speaker's way of presenting the embedded proposition P' (2001: 70-1). The way to contradict you is to deny P, not to deny that you believe that P." (Cassam, 2014, p. 145). Inferentialism says that if 'I believe that *p*' is a genuine statement of what you believe, if you

are questioned as to whether you believe that *p*, you come to your conclusion by a conscious or unconscious inference from *p*. This brings out that, for Cassam, although some self-knowledge is inferentially grounded on facts about one's behaviour, some is not. Paradigmatically, self-knowledge of one's beliefs is grounded inferentially in the manner described by Byrne: even in cases where one successfully follows BEL, the inferential basis of such self-knowledge is not facts about one's behaviour (expect in those cases where the beliefs in question are beliefs about one's behaviour, of course). Cassam's conclusion is that by a combination of arguments from elimination of alternatives, the examples of inferentialism providing a parsimonious explanation and the experimental evidence, we should take it that the moderate inferentialism he sets out is

> "…a coherent, intuitively plausible, and well-supported alternative to the view that intentional self-knowledge is, if not impossible, then based on inner observation or based on nothing." (Cassam, 2014, p. 148)

Cassam's moderate inferentialism amounts to the idea that self-knowledge *can* be based on inference from behaviour, or psychological cues, but an account of self-knowledge is not exhausted by the suggestion that self-knowledge is based on inference. Cassam sees this moderate inferentialism as the most plausible account of self-knowledge on offer. Throughout the first part of *Self-Knowledge for Humans*, Cassam takes himself to have ruled out all non-inferential accounts of self-knowledge. I suggest that this is perhaps too strong a position, and that there is a Rationalist alternative in the offing which is neither an inner sense nor an Inferentialist view, and which does not lead to 'cognitively insubstantial' self-knowledge, and further does not presuppose that the knower is *homo philosophicus* or fall into the over-intellectualisation objection. I will argue for this view in chapters three and four, but before going further we should survey the alternative to inferentialism.

1.3.3. Rationalist Accounts

Rationalist Accounts of Transparency aim to make sense of the link between the answer to the question 'is it the case that *p*?' and the answer to the question 'do you believe that *p*?' by appealing to rational or critical capacities, broadly characterised by Burge's notion of *Critical Reasoning*:

> "Critical reasoning is reasoning that involves an ability to recognize and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. As a critical reasoner, one not only reasons. One recognizes reasons as reasons." (Burge, 1996, p. 98)

One way to make sense of Burge's idea is to understand Evans' remark as pointing to a *constitutive* account of self-knowledge. Constitutive accounts in general, however, do not take themselves to be

developing Evans' remark as such. Rather, they follow a broadly Wittgensteinian motivation[15] to make sense of an apparent conceptual connection between first and second order attitudes (paradigmatically avowals of sensations like 'I am in pain' and 'I believe I am in pain') A constitutive account suggests that there is an *a priori* conceptual connection between a subject's first-order mental states (e.g., beliefs, desires, intentions) and their second-order mental states. As Marcus and Schwenkler put it

> "**Constitutivist** theories hold that we can have non-empirical knowledge of our beliefs because to take oneself to believe something is, at least in the right conditions, also to believe it." (Marcus and Schwenkler, 2018, p. 15)

Of course, such a characterisation is still somewhat loose. Coliva (2016) gives the following characterisation of the Constitutive thesis as a biconditional:

> "**Constitutive Thesis**: Given C, one believes/desires/intends that P/to Φ iff one believes (or judges) that one believes/desires/intends that P/to Φ." (Coliva, 2016, p. 164)

The business of the constitutive account, on these terms, is elucidating the C-conditions under which the constitutive thesis holds and giving an account of the order of explanation of the biconditional; that is, an account which explains whether and why the left-to-right direction of explanation takes priority[16], the right-to-left[17], or whether there is no priority[18] in the direction of explanation.

A detailed treatment of these varieties of constitutivism is beyond the scope of this work, given the focus on Evans' Transparency Remark. However, Marcus and Schwenkler (2018) note the similarity between constitutivism and Evans' remark, and suggest that one way of understanding Evans' remark is by suggesting a constitutive relation between the answer to the question 'is it the case that *p*?' and the answer to the question 'do you believe that *p*?', such that answering the question of what is the case constitutes giving an answer to the question of belief. This would be, on Coliva's terms, a form of left-to-right constitutive account. Marcus and Schwenkler suggest, however, that a constitutive account of Evans' remark is unsatisfying as constitutive accounts "…seem to require seeing doxastic self-attribution as entirely groundless, as the product of a brute disposition or something along these lines." (Marcus and Schwenkler, 2018, p. 15) While I do not think this is a

[15] Of course, Evans' account shares some of this motivation in the rejection of 'inner sense' notions of introspection, as noted by Byrne:

> "To find out that one sees a blue mug, one does not turn one's attention inward to the contents of one's own mind— Moore's and Wittgenstein's remarks suggest either that there is no such procedure or, if there is, it is not necessary. Rather, one turns one's attention outward, to the mug in one's environment. This insight—if that is what it is—was first clearly expressed by Evans." (Byrne, 2018, p. 3)

[16] See e.g. Shoemaker(1996)

[17] See e.g. Wright (2001)

[18] See e.g. Bilgrami (2012)

devastating objection, a more pressing objection is that saying that one answer is constituted by the other is no more explanatory than saying that one answer is transparent to the other; so a significant amount of explanatory work in what the relation of constitution amounts to remains to be done. It should be further noted at this point that the constitutivist Marcus and Schwenkler engage with is of the 'right-left' variety, one who thinks that believing that one believes that *p* ensures or settles it that one believes that *p*. Nevertheless, I think their criticism has teeth against any view of self-knowledge which takes belief to be a brute disposition. Such an account seems like an unsatisfying explanation of self-knowledge.

Marcus and Schwenker (2018) offer what they take to be an alternative reading of the same sort of thought (following Marcus (2016)) in claiming that belief is *self-conscious*, that

> "…it is in the nature of belief to be self-conscious, and that because of this a person does not need to rely on introspection or any other empirical process as a source of knowledge of her beliefs." (Marcus and Schwenkler, 2018, p. 14)

This view in many ways tracks the same sort of insight that motivates the constitutive account. On the constitutive view, the answer to the question 'do you believe that *p*?' is (at least partially) constituted by the answer to the question 'is it the case that *p*?' On the view of Transparency as a consequence of the self-consciousness of belief, it is simply a feature of answering the question 'is it the case that *p*?' that the subject is in a position to answer the question 'do you believe that *p*?' It might seem as if there is little water between the constitutive notion of Transparency and the self-conscious notion of Transparency. However, Marcus and Schwenkler consider the structure of explanation to be a significant difference. They put it like so:

> "The thought is not that (as for the Constitutivist) the thinker simply finds herself believing that she believes that p and, in virtue of this fact, believes that p. Rather, according to the Self-Conscious Conception a person is ordinarily in a position to grasp what she believes simply in virtue of believing it, as it is part of what it is to view a proposition with the belief-attitude that one thereby knows oneself to so view it." (Marcus and Schwenkler, 2018, p. 16)

On the Self-Conscious Conception, the insight captured by the Transparency Remark is a consequence of the nature of belief and what it is to be a believer. As Marcus and Schwenkler put the Self-Conscious Conception,

> "…a person is ordinarily in a position to grasp what she believes simply in virtue of believing it, as it is part of what it is to view a proposition with the belief-attitude that one thereby knows oneself to so view it." (Marcus and Schwenkler, 2018, p. 16)

In answering the question 'is it the case that *p*' a subject is *inter-alia* in a position to answer a question regarding her belief, because the answer to the question of what is the case entails that she has a self-conscious belief regarding what is the case, and thus knows herself to believe it. The

constitutivist holds that a biconditional holds between a first order belief and a second order belief (under certain conditions), that is, if a subject believes that *p*, she believes that she believes *p* (and vice-versa). The Self-Conscious conception makes the same claim; it suggests that believing that *p* puts one in a position to believe (and *inter alia* know) what one believes. The Self-conscious conception is a variety of constitutivist account, but not of the left-right form that Marcus and Schwenkler take to be the constitutive view. If believing *p* puts one in a position to know that one believes *p*, this would seem to be a form of left-to-right constitutivism. The work of avoiding the objection that the self-knowledge provided by the constitutive account is given by nothing more than a brute disposition is done by talk of the self-consciousness of belief or by Evans' talk of a 'procedure'. Of course, these ideas must be substantiated to answer the objection, and that is the business of a Transparency account.

Moran (2001) aims to give a thoroughgoing Rationalist (and self-conscious) conception of the Transparency Remark by tying Evans' notion to *agency*. The concern of *Authority and Estrangement* is to explain how self-knowledge can be both *Authoritative* and *Immediate* (i.e., groundless), and the Transparency remark gives Moran a starting point for doing this. We can see Moran's rationalism (in connection with Burge above) in the following understanding of the Transparency Remark:

> "So, rather than reducibility or indistinguishability, the relation of transparency these writers are pointing toward concerns a claim about how a set of questions is to be answered, what sorts of reasons are to be taken as relevant." (Moran, 2001, p. 62)

The focus of Moran's text is the idea that answering the world-directed question puts one in a position to answer the question of belief. His suggestion is that one is in a position to answer the question of belief because one *deliberates* or *makes up one's mind* regarding the world-directed question. Deliberation puts one in a position to answer the belief question, as to make up one's mind about *p* is to form a belief that *p*, so one is in a position to answer the question 'do you believe that *p*?'. This deliberative stance toward the question 'is it the case that *p*?' pulls the question of belief and the question of how things are together for the subject. As Moran puts it:

> "With respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question of the truth of P." (Moran, 2001, pp. 62-63)

It is worth noting at this point the link between Moran's proposal and the Self-Consciousness Conception suggested by Marcus and Schwenkler above. Moran's proposal makes significant use of the idea of *self-constitution* as a feature of self-knowledge, and the unpacking of Moran's deliberative proposal has much in common with the constitutive reading of the Self-Consciousness Conception suggested above. Indeed, we might think of Moran's project is to give a conception of

substantive self-knowledge which respects the insight of the Self-Consciousness Conception while explaining Authority and Groundlessness. The respect for the Self-Consciousness Conception of self-knowledge leads to an important secondary strand in Moran's discussion; Moran suggests his account does not explain all self-knowledge, and it does not have designs on doing so – he restricts his account to self-knowledge of attitudes. However, he suggests that the sort of self-knowledge in question is *fundamental* self-knowledge, although it is not clear from his text what this fundamentality amounts to. Boyle (2009) presents a substantial discussion of this fundamentality point, and suggests that the capacity to deliberate is fundamental to being a self-knower at all. Boyle's discussion of fundamentality (in response to the Scope objection to Transparency, see section 1.4.4. below) forms chapter four of this work and is discussed in considerable detail there.

Boyle (2011) attempts to give an alternative Rationalist account which respects both the Transparency Remark and Moran's insights about deliberation over one's beliefs. Boyle characterises his own approach as a *reflective* approach to Transparency:

> "Instead of thinking of the subject as making an inference from *P* to *I believe P,* he can think of the subject as taking a different sort of step, from *believing* P to *reflectively judging* (i.e., consciously thinking to himself): I believe P. The step, in other words, will not be an inferential transition between contents, but a coming to explicit acknowledgment of a *condition* of which one is already tacitly aware. The traditional philosophical term for this sort of cognitive step is "reflection," so I will call this a reflective approach to explaining transparency." (Boyle, 2011, p. 5)

The key insight of this discussion is the idea that the transparency of belief is not a matter of inferring or concluding some fact about myself from a fact about the world, but is rather a shift of attention, from the world at large to one's engagement with it. An important consequence of this is the idea that Boyle's approach does not explain how we acquire doxastic self-knowledge, but rather says something about the nature of being a believer. In this sense, it follows Marcus and Schwenkler's Self-Consciousness Conception of Transparency. Boyle's account…

> "… treats the following as a basic, irreducible fact about believing as it occurs in a creature capable of reflection: a subject in this condition is such as to be tacitly cognizant of being in this condition. Hence, in the normal and basic case, believing P and knowing oneself to believe P are not two cognitive states; they are two aspects of one cognitive state – the state, as we might put it, of knowingly believing P." (Boyle, 2011, p. 6)

On Boyle's view here, the actualization of the power one actualizes in believing *p* is the very same actualization of a power as actualized in knowing that one believes that *p*. Boyle's account here takes believing to be self-conscious, insofar as believing that *p* simply *is* tacitly knowing that you believe that *p* – they are one and the same cognitive state. An important upshot of this is in the sort of explanation of self-knowledge Boyle suggests an account of this form offers. He suggests we

distinguish *epistemic* accounts of self-knowledge from *metaphysical* accounts. Epistemic accounts (according to Boyle) seek to explain how the relation between being in a mental state M and believing that one is in a mental state M is such that the latter state can amount to knowledge of the former state. Boyle's account rejects this problematic – being in a mental state M simply is tacitly knowing one is in a mental state M (at least in the case of belief). It is a constitutive feature of those states we bring under the rubric of self-knowledge. The *metaphysical* account aims to explain self-knowledge by giving an account of the nature of the various mental states of interest and why those states obtaining entails tacit knowledge of them. Boyle suggests in this account that belief is such a state, and even if belief is the only such state, it would be of central importance to doxastic self-knowledge in general.

Finkelstein (2012) offers a novel reading of the Transparency Remark in response to several objections to Moran's deliberative account. Finkelstein's suggestion is that the proper way to understand the Transparency Remark is by realising that we learn to use the sentences of the form '*p*' and 'I believe *p*' interchangeably.

> "Thus, having just looked out your window at a fast-approaching bank of black clouds, you might say *either* "It's about to rain", or "I believe it's about to rain." The latter assertion calls for no more inward observation than the former." (Finkelstein, 2012, pp. 113-114)

Finkelstein's reading of this interchangeability in sentence use is that if 'I believe it's about to rain' is an expression of the speaker's mind, then 'it's about to rain', if it is interchangeable with the 'I believe…' sentence, equally expresses the speaker's mind, that we should

> "…understand the relation between your belief and your self-ascription of belief as akin to that between your sadness and your crying – so: not as depending on inward detection, but as an expression, a manifestation, of your psychological condition." (Finkelstein, 2012, p. 115)

There is, I think, something right in Finkelstein's thought here, but I suggest that it is related to the lack of exercise of a further epistemic capacity in the assertion of belief over the assertion of how things are.[19][20]

Fernandez (2013) aims to develop a broadly naturalist understanding of self-knowledge through the Transparency Remark. Fernandez's explanatory goals are to explain what he calls *special access*, which suggests that self-knowledge is had neither on the basis of reasoning nor on the basis of behavioural evidence, and *strong access*, where, in the normal run of things, a subject *S* is more justified in her own belief attributions than in her attributions of belief to another. Fernandez takes

[19] We will return to Finkelstein's proposal and how it differs from the positive proposal of this thesis in section 6.1.2.
[20] This is discussed in detail in chapter three

these two properties to be constitutive of Privileged Access. In developing the background of his account, Fernandez introduces a further useful distinction, which tracks the distinction suggested by Boyle between *epistemic* and *metaphysical* accounts of self-knowledge. Fernandez suggests we can separate explanations of self-knowledge into the *Doxastic* and *Non-Doxastic*:

> "The doxastic versus non-doxastic distinction is a distinction about the explanandum being targeted by an investigation of self-knowledge. Doxastic investigations try to account for facts about higher-order beliefs whereas non-doxastic investigations do not try to account for such facts." (Fernandez, 2013, p. 10)

This distinction matches Boyle in the sense that a Doxastic account of self-knowledge must explicitly be what Boyle would call an Epistemic account, but a Non-Doxastic account need not be.[21] Fernandez introduces two further useful distinctions (which cut across each other). Firstly, between Causal and Non-Causal accounts of self-knowledge, where the distinction is between those accounts which explain privileged access in terms of a causal relation between a subject S's mental states and her beliefs about those states, and those accounts that propose an alternative relation. Secondly, the distinction between Reasons and No-Reasons accounts. Reasons accounts suggest that "…our special, strong access to our mental states requires having reasons of some kind for believing that we are in those states." (Fernandez, 2013, p. 26), and the No-Reasons account does not. It is worth noting that as presented, the Causal/Non-Causal and Reasons/No-reasons distinctions both fall on the Epistemic side of Boyle's Epistemic/Metaphysical distinction. Fernandez carves up the territory of self-knowledge somewhat using these distinctions, but for our purposes, we need only note that they are a useful way of carving up the discussion of explorations of the Transparency Remark.

Fernandez's account focuses on the role the grounds of a belief play in the explanation of that belief. The basic ideas is that "…we self-attribute beliefs on the basis of our grounds for those beliefs." (Fernandez, 2013, p. 49) So the subject would attribute her belief that *p* to herself on the basis of her grounds for her belief that *p*. Fernandez develops the Transparency Remark by suggesting that the self-attribution of a belief 'bypasses' the belief being self-attributed:

> "The basic idea, then, is that we self-attribute beliefs on the basis of our grounds for those beliefs. I will use 'bypass' to refer to the procedure whereby a self-attribution of a belief is formed on the basis of grounds that the subject has for the self-attributed belief." (Fernandez, 2013, p. 49)

This view is formulated by Fernandez as follows:

> "*The bypass view (Belief)*

[21] Note that such an account still could be an Epistemic account. Boyle's own account would by this distinction be a Non-Doxastic Metaphysical account of self-knowledge.

> For any proposition P and subject S:
>
> Normally, if S believes that she believes that P, then there is a state E such that
>
> (a) S's (higher-order) belief has been formed on the basis of E.
>
> (b) E constitutes grounds for the belief that P in S." (Fernandez, 2013, p. 49)

Fernandez points to the Transparency Remark as motivation for the idea that the self-attribution of beliefs 'bypasses those beliefs'. He credits Evans with the idea that when I am asked 'do you believe there will be a third world war?' my self-attribution of belief in answering this question is not done on the basis of the belief in question (my belief regarding a third world war), but rather on the basis of the grounds for that belief (the grounds I have for believing there will be a third world war).

A key idea we should be aware of for this view is the importance of the *basing relation* to the view. The belief which would amount to self-knowledge (the higher-order belief) is formed on the basis of the state E, which is the grounds or basis of the first order belief that P. There are two necessary conditions for the basing relation to obtain according to Fernandez:

> "For any subject S, proposition P and state E:
>
> 1. If S forms the belief that P on the basis of E, then S believes that P because she is in E.
>
> 2. If S forms the belief that P on the basis of E, then S is disposed to believe that she is in E (provided that she reflects on why she is forming her belief, and she has the appropriate conceptual repertoire)." (Fernandez, 2013, p. 41)

Condition (1) tells us that if the subject *S* forms a belief on the basis of some state, then the fact that she is in such a state is the cause of her having the belief. This suggests that any account which aims to suggest the link between the first and higher order belief states is in terms of a basing relation is a causal account in Fernandez's taxonomy. Condition (2) suggests that if the subject forms a belief on the basis of some state, she is in a position to believe she is in the state that is the basis of the belief she has formed.

Fernandez's account places the grounds for belief in centre stage. This centrality of the grounds for one's higher order beliefs is an extension of Fernandez' focus on the justification of one's self-attributed beliefs. Fernandez's explanatory goal is to account for privileged access, and the Bypass view he has put forward aims to do that by explaining (what he calls) Strong access and Special access.

The explanation of Special access relies on the following idea:

> "[F]orming a belief on the basis of some state does not require believing that one occupies that state, and it does not require believing that, if one is in that state, then the content of the belief being formed is likely to be the case. It is just a matter of, as it were, trusting the relevant state, or taking it at face value." (Fernandez, 2013, p. 56)

Forming my first order belief that *p* on the basis of my perceptual experience *p* does not require that I form a further belief that my perceptual apparatus is reliable, or a belief that I seem to perceive *p*. All I need to do is accept the way my experience presents the world to me. Fernandez suggests this is no different for higher order beliefs:

> "In order to form my belief that I believe that there is an apple in front of me, I do not need to believe that I seem to perceive one, and I do not need to believe that I usually believe that there is an apple in front of me when I seem to perceive one. Thus, I do not need to resort to behavioural evidence or, for that matter, any other source of information to arrive at those two beliefs. The reason why I do not is that I do not need to use those beliefs as premisses in an inference towards the conclusion that I believe that there is an apple in front of me. I just need to take my perceptual experiences at face value." (Fernandez, 2013, pp. 56-7)

We can see that this meets Fernandez's restriction on Special access; a subject who comes to her higher-order belief by the Bypass method will not have come to that belief on the basis of behavioural evidence, and neither will she have reasoned her way there, via an inference or other reasoning pattern.

The explanation of Strong access is likewise quite straightforward. Fernandez suggests that by self-attributing beliefs via the Bypass method, a subject's belief attribution is less liable to error than those beliefs attributed to others using a different method:

> "Consider the scenario in which, unbeknownst to me, my perceptual experiences are often wrong. […] In itself, that would not render my bypassing self-attributions of perceptual beliefs unjustified. For the correlation between the world and my perceptual experiences that justifies my perceptual beliefs is independent from the correlation between those experiences and the perceptual beliefs that they generate. And my justification for my self-attributions of perceptual beliefs relies on the latter correlation." (Fernandez, 2013, p. 58)

That my self-attributions of belief by the Bypass method are not prone to the same error as attributions of belief made by another method (say an inference from behavioural evidence) is taken by Fernandez to secure Strong access, the idea that the self-attributions of belief in question are in better shape epistemically than attributions to others by an alternate method. By explaining Strong access and Special access, Fernandez takes himself to have secured an account of the privileged access of belief.

Fernandez's account also aims to generalise the Bypass model into a discussion of the transparency of desires, although given the brief treatment of Fernandez's account here, we will not discuss this in any detail. Fernandez further aims to explain both the phenomenon of self-deception and that of thought-insertion — the idea that some thoughts might phenomenologically feel to the subject like they do not belong to the subject having them. Of the Rationalist approaches discussed thus far, Fernandez's approach both appears to stay closest to Evans' original remark by focussing on the

grounds or basis of the beliefs in question, and is closest to the Inferentialist programme[22], both in terms of some of his aims in explaining self-knowledge in terms of the Transparency Remark (e.g. by asking what it is that the inappropriateness of Moore's Paradox actually tracks), and in terms of the mode of explanation, by focussing on the basing relation.

O'Brien (2005) aims to deliver an account of how a subject might have knowledge of her *assertoric acts of mind* inspired by the Transparency Remark. These assertoric acts of mind are, O'Brien suggests, those of judging that *p*, denying that *p*, questioning whether p and doubting that *p*. (O'Brien, 2005, p. 581) O'Brien's account bears many similarities to Moran's deliberative account discussed above, but O'Brien is more explicit about the role of rational agency in the provision of the *warrant* for the knowledge a subject has of her assertoric acts of mind[23]. The approach to explaining the warrant that a subject has for the knowledge of her judgements draws on the idea of a *rational entitlement* to the warrant for the judgement 'I φ that *p*' (where φ is an assertoric act of mind). O'Brien's suggestion is that

> "The essence of the agency theory is that the rational connection between the pre-suppositions of rational agency, and the self-ascriptions we are concerned with, is of a kind that automatically makes those self-ascriptions ones to which the subject is entitled in the absence of evidence or reasons." (O'Brien, 2005, p. 591)

The thought is that there is something special about the nature of rational agency which secures the warrant for the judgements in question, and the task is to show what that amounts to. O'Brien notes that it is not hard to see the transition from judging *p* to judging 'I judge that *p*' as a reliable one; "…on such a transition, the means by which the self-ascription is reached is that which is self-ascribed, [so] self-ascriptions made [on this] basis will track the truth." (O'Brien, 2005, p. 591) [clarifications mine]. Rather, the puzzle is how can such a transition be rational. The irrationality of the transition echoes the inferential inappropriateness of the Doxastic Schema in Gallois; how can it be a rational transition to go from the judgement that *p* to the judgement that 'I believe that *p*'[24]. On the face of it, there is nothing relating the content of the judgement that *p*, which is concerned with how things are with the world, and the content of the judgement 'I believe that *p*', which concerns how things are with the subject. O'Brien suggests that the gap between the judgement that *p* and

[22] Note that Fernandez's account could not (at least on the face of it) be subsumed into an inferentialist account as Special access as he presents it denies that self-knowledge could be the result of a reasoning process.

[23] From here on when discussing O'Brien's proposal, I will talk of judgements, rather than 'assertoric acts of mind'. Unless explicitly noted, the reader can substitute any of the assertoric acts of mind O'Brien notes in for judgement.

[24] This, along with the inappropriateness of the Doxastic Schema, is a way of putting a more general puzzle surrounding the Transparency Remark – how can answering a question about the world tell me *anything* about myself? This is a standard objection to Transparency views and will be detailed in more depth in section 1.4.1. of this chapter.

the judgement 'I believe that *p*' is closed when we consider the judgement that *p* as a produce of rational agency; "…that is, in the context of the subject determining what attitude she will adopt by a consideration of what is true." (O'Brien, 2005, pp. 591-592). This view shares much in common with Moran, centrally for O'Brien's purposes

> "…a subject who self-ascribes an attitude, guided by her consideration of what is true, is entitled to take the attitude as being an attitude of φ-ing, because φ-ing is the attitude, which she has practical knowledge of as a possibility and which her consideration of what is true immediately led her to adopt." (O'Brien, 2005, p. 592)

The subject's practical knowledge of what assertoric attitude is a possibility for her to adopt given her first order judgement (the judgement that *p*) is what allows the transition between the first order judgement and the second order judgement to be a rational one, since her practical knowledge entitles her (in the sense of epistemic entitlement), in virtue of her rational agency, to choose to adopt one of the attitudes in question. O'Brien presents the following considerations to motivate such a position:

> "1. Being a rational agent means determining ones attitudes on the basis of reason.
>
> 2. Determining one's assertoric attitudes on the basis of reason means determining ones attitudes by a consideration of what is true.
>
> 3. Determining one's attitudes by a consideration of what is true presupposes that judging, denying, questioning or doubting are options one can immediately implement in a given instance on the basis of such a consideration
>
> 4. It is only one's own attitudes that one can immediately form or change on the basis of such a consideration.
>
> 5. The force of the attitude, with respect to P, that one immediately forms or changes on the basis on a consideration of what is true is determined by one's conclusion with respect to the truth of P i.e. by whether one's conclusion is P, not-P, P is unsettled or unlikely.
>
> 6. A subject exercising reason over her thought must have a practical knowledge of her options to judge, deny or doubt as (a) things can be done and (b) as things that can be done by her." (O'Brien, 2005, pp. 592-593)

In this way, O'Brien presents us with a reading of the Transparency Remark with reasons and agency at the fore. This reading shares some commonalities with Moran (and to some extent Boyle), with a notable difference in explanatory goal. Moran's target is an explanation of the authority and immediacy of some sorts of self-knowledge, but O'Brien's target is the justificatory status of certain sorts of judgements. Moran is explicitly tackling a subject's beliefs, whereas O'Brien restricts her accounts to assertoric attitudes, and is explicit about *not* tackling beliefs. Further, there is a central puzzle regarding O'Brien's account. O'Brien takes for granted that the subject has 'practical knowledge of her options'. But if this is right, then it seems that when an agent determines her attitudes, she does so on the basis of this practical knowledge, and as such does so *intentionally*, and

as such does so *in the knowledge that she is doing it*. If so, there is no explanation of self-knowledge in the offing here. Further we might wonder if 'practical knowledge of her options' amounts to self-knowledge itself, adding a still further dimension of circularity. Without a fuller understanding of these ideas, we cannot rule out this circularity. Nevertheless, the idea that there is a significant connection between self-knowledge and rational agency has much to recommend it, and something of this idea is preserved in the Simple Account presented in chapters three and four.

1.4. Standard Objections to Transparency

Any theory of self-knowledge aiming to develop the Transparency Remark must engage with and overcome or dissolve several standard objections. In this section I will give a brief summary of a standard form of each of these objections (different explications of the Transparency Remark will be vulnerable to these objections in differing ways, of course), but I will not offer a substantive dissolution or response to them, instead I will suggest how they might be tackled by some of the views discussed, where that is appropriate. In particular, I will respond to these objections in chapter three when discussing my own positive view of the Transparency Remark.

1.4.1. The Puzzle of Transparency

The primary objection any theorist trying to make sense of Evans' remark may have is what Byrne calls The Puzzle of Transparency. Byrne's articulation of this objection is formulated in terms of his own inferential account, but there is a general form of this puzzle that can be specialised into either an Inferentialist or a Rationalist guise. In either guise, the puzzle trades on the following thought: How can examining how things are with the world *possibly* tell me anything about myself? That is, how can answering a question about whether or not there will be a third world war possibly tell me whether or not I believe there will be? The Inferentialist guise of this objection asks how the inference from *p* to believing that one believes that *p* can possibly be a good inference, and if the inference is not a good one, how can it be knowledge-producing? Byrne, for example, suggests the beliefs produced by the transparent inference are *safe* (i.e., could not easily be wrong) and as safety is a sufficient condition on knowledge, the inference is knowledge producing. Any Inferentialist account will need to contend with this form of the puzzle, and what Boyle calls the 'mad inference' version of the same point: even if *p* is true,

> " …the truth or otherwise of *p* has no tendency to show that I believe it What would support my conclusion, of course, is the fact that I, the maker of this inference, accept the premise that *p*. But to represent that as my basis would be to presuppose that I already know my own mind on the matter, and that would undermine Byrne's account." (Boyle, 2011, p. 8)

Of course, Boyle's objection here is levelled at Byrne, but a more general objection to the Inferentialist programme is in the offing – the idea is that the basis of the inference must exhibit some support relation with the consequent of the inference, and that support relation must not

amount to a piece of self-knowledge. This is a form of the general Puzzle of Transparency to which all Inferentialist accounts must be sensitive.

The Rationalist version of the puzzle is captured clearly by O'Brien (2005):

> "[F]rom the subject's perspective the transition seems to cross a gap between radically different and unrelated contents. The subject judges, say, that the sky is blue. How is it in any way rational for her to judge 'I judge [that] the sky is blue' on that basis. How is it that the transition, from a content about the sky, to a content which involves the first person and her attitudes, is rational for the subject?" (p. 591) [clarification mine]

That is, (in O'Brien's terms), how can the content of the judgement *p* license a rational transition to the judgement 'I judge that *p*'?[25] The Rationalist must explain how the transition the Transparency Remark articulates is a rational one. O'Brien and Moran both appeal to agency to explain how the agent can be credited as rational. But the Rationalist has a resource not available to the Inferentialist in answering the puzzle – the Rationalist can deny that there is a transition at all. Boyle moves toward this solution, and the positive account I suggest in chapter four gives a detailed response to this objection which is influenced by this idea.

1.4.2. The Anti Luminosity Argument

Timothy Williamson's Anti Luminosity Argument (Williamson, 2000, ch. 4) generates a significant problem for any attempt to develop the Transparency Remark into an account of self-knowledge. The conclusion of the Anti-Luminosity argument is that no interesting states are *luminous*, where a luminous state is a state *X*, where necessarily, if you are in *X*, you are in a position to know you are in *X*. Evans' remark suggests that by answering the question 'is it the case that *p*?' you are in a position to answer the question 'do I believe that *p*?' The target of the explanation of the Transparency Remark is a subject's self-knowledge. So, what the friend of Transparency is after is the result that believing *p* (answering the 'is it the case that *p*?' question) puts one in a position to *know* that one believes that *p* This is the denial of the anti-luminosity claim (i.e., Transparency is the affirmation of the claim that belief is luminous). The friend of Transparency must show how their account of belief (and of any state they wish to explain as self-knowledge) is not threatened by Williamson's argument, and this, I think, is a challenge that is not engaged with well in the literature. Rather, there is often what appears to be a dialectical impasse, where intuitions and argument support Transparency and the Anti-Luminosity Argument denies it. In section 3.5.1. I will aim to give a Rationalist response to Anti-Luminosity which preserves a central idea Williamson takes to be a

[25] The Transparency Remark is generally cashed out in terms of belief, and thus the puzzle is generally formulated in those terms. I have used O'Brien's version which centres on judgement, but I see no reason why it could not be reformulated in terms of belief. The key point is that the transition does not seem to be licensed merely by an appeal to rationality, since it seems (prima facie) that there is nothing rational about concluding 'I believe *p*' on the basis of '*p*'.

motivation for the Anti-Luminosity argument, but also show the argument to get no purchase on a properly understood articulation of the Transparency Remark.

1.4.3. Over-Intellectualisation

Transparency accounts can seem to 'over-intellectualise' our everyday self-knowledge. This puzzle arises for both Inferentialist and Rationalist accounts but seems more pressing in the Rationalist case. For the Inferentialist, who claims that our self-knowledge is the result of an inference from some premiss to the conclusion that we know what we believe, the charge of over-intellectualisation is that we simply don't make these inferences all the time (or even in the normal case) of knowing our own mental states. This over-intellectualisation worry also highlights that inferentialism seems to undermine the immediacy or groundlessness of self-knowledge; if self-knowledge is the result of an inference, in what way is it immediate or groundless? The Inferentialist can respond to the charge that inferentialism denies immediacy by biting the bullet and saying the feeling of immediacy we have in self-knowledge is just that, a feeling (indeed, a mere feeling) and that the inference is made so quickly as to give that feeling. The Inferentialist might also claim that the inference is not conscious, in the manner of Cassam above, or could even take the stronger position that the talk of 'inference' is not the personal level inference as we would understand it, but is instead a sub-personal process, and talk of inference is purely heuristic[26]. Regardless, the Inferentialist is in a strong position to disarm this objection.

The Rationalist explications of the Transparency Remark also find themselves vulnerable to the charge of over-intellectualisation. For the Rationalist, the charge is that however the rational relation between the answer to the world-directed question and the answer to the question of belief is understood, it will be in a way which does not do justice to the idea that self-knowledge is often effortless and immediate. The over-intellectualisation objection says of the deliberative account that in very many cases we don't 'make up our minds' as to how things are before coming to a piece of self-knowledge, and as such, the deliberative account just gets the phenomenon wrong. The objection suggests that self-knowledge does not involve the sort of rational transition between contents the Rationalist conception suggests, as any rational transition would undermine the immediacy of self-knowledge, or minimally, characterises what goes on in the normal case of self-knowledge incorrectly, even if the Rationalist does correctly characterise at least some self-knowledge correctly. The Rationalist could respond in the manner of Moran, by suggesting that even

[26] There is some sense in which taking this route means whatever explanation of self-knowledge the inferentialist provides, it is *not* an explanation which builds on the Transparency Remark – the Transparency Remark as Evans has it focusses on the answering of questions, a *personal level* activity. Any account which understands Transparency as a sub personal process understands Transparency in a way deeply divergent from Evans.

if the over-intellectualisation objection is correct, the aim is not to provide a unified account of self-knowledge, and the self-knowledge explained by the Rationalist account is in some sense fundamental or central to a more general account of self-knowledge. I will suggest in chapter four that this response is too concessive, and that there is a unified Rationalist account which is both fundamental in the sense Moran wants and satisfies the Objection from Over-Intellectualisation. This leads us to the final, and I suggest most pressing worry for the friend of Transparency.

1.4.4. The Objection from Scope

The final of the central objections to Transparency is the objection that the scope of the account delivered is too narrow. This objection relies on the premiss that any account of self-knowledge should be *general* (or *unified*); a general account explains all self-knowledge in a single explanatory swoop. The alternative position suggests that self-knowledge of belief is explained differently from self-knowledge of desires, which in turn is explained differently from self-knowledge of sensations, and so on. The Rationalist version of this objection is developed and tackled in detail in chapter four, where I give a formula for Rationalist account of the Transparency Remark which is completely general, and as such not vulnerable to the Objection from Scope. As the solution I suggest is tied intimately to the particular development of the Transparency Remark in chapters three and four, it not available to the Inferentialist, but Byrne at least specifically aims to give a unified account of self-knowledge and as such is not vulnerable to this objection.

1.5. Concluding Remarks to Chapter 1

This chapter has achieved three central goals which will condition the approach the rest of the thesis will take. First, it has suggested what we want out of an account of self-knowledge – we want either an explanation of privileged access (through an explanation of the two aspects it can be decomposed into, *groundlessness* and *authority*). Further, we want a *general* account that respects the Objection from Scope. Second, it has partitioned the literature along Rationalist and Inferentialist lines and has given an overview of the development of accounts which build on Evans' Transparency Remark with that partition in mind. Third, it has outlined the central objections to a Transparency account, objections that any satisfactory account of Evans' remark must either answer or dissolve.

With these three central goals in mind, and an overview of the literature in place, we are now in a position to examine the background and motivations of Evans' account to draw out what I take to be the central idea expressed by the Transparency Remark.

2. Evans' Transparency Remark

As discussed in chapter one, Gareth Evans, in *The Varieties of Reference* (Evans, 1982), develops a version of what has become known as a *Transparency* account of self-knowledge. There, I gave a limited exegesis of Evans' account to allow for a discussion of how Evans' work has since been developed. In this chapter we return to Evans' own development of the Transparency Remark, and look in particular at the background thought, drawn from Strawson and Kant, which motivates and constrains Evans' account. In doing so, I aim to expose the beautifully simple idea that lies at the heart of the Transparency Remark.

The central expression of Evans' Transparency account is taken to be the following remark from *Varieties*:

> "If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting in to operation whatever procedure I have for answering the question whether *p*." (Evans, 1982, p. 225)

Evans' great insight here is that to answer a question regarding herself, (i.e., whether she believes there will be a third world war, a question targeting her own beliefs), a subject does not 'look inward' and seek to examine her beliefs, rather she engages in whatever procedure would enable her to answer the question whether something is the case (i.e., answer the question 'will there be a third world war?'). Evans' Transparency account suggests that there is some important sense in which the question 'do you believe that *p*?' is *transparent* to the question 'is it the case that *p*?'[27]. To understand Evans' account of self-knowledge, to understand what it is for the questions to be transparent, we will examine the motivation for Evans' explication of the Transparency remark, and in doing so, develop constraints which arise from Strawson's Kant[28].

2.1. Motivating a Transparency Account

We might think that the first and most obvious motivation for a Transparency-based account of self-knowledge is that Evans' insight seems *intuitively plausible*. There is something congenial in the thought that to know what I believe I need only answer a question regarding how things are, when this thought is put in the way Evans does in the Transparency Remark. But this is no substantive motivation for a Transparency account. Rather, to motivate an account of self-knowledge which

[27] Alternatively, 'whether *p*?'

[28] In particular, Strawson's *The Bounds of Sense* (Strawson, 1966). Similar thoughts can be seen in *Individuals* (Strawson, 2003), but this chapter draws on Strawson's explicit Kantian exegesis and exploration.

explicates or develops Evans' remark, we should look to the Puzzle of Self-Knowledge from section 1.1.

The Transparency Remark gives a relatively straightforward explanation of the apparent groundlessness of our beliefs. When an interlocutor asks us 'do you believe that *p*', we come to assert or affirm our belief in *p* by "…putting in to operation whatever procedure [we] have for answering the question whether *p*." (Evans, 1982, p. 225). Thus, self-knowledge appears groundless; to answer the question of belief one puts into operation whatever procedure one puts into place to answer the world directed question. The subject need do nothing more to answer the question of belief than answer the question of what is the case. Her knowledge of her belief is groundless in this sense. Of course, much depends on exactly how 'whatever procedure I have for answering' is cashed out, but the idea is that engaging in that procedure in the face of the interlocutor does provide some sort of direct or immediate or groundless awareness of our belief regarding *p*[29]. Although an explanation of groundlessness seems to have been attained, at least in the case of self-knowledge of beliefs, an explanation of the authority self-knowledge exhibits seems not to follow quite so easily. Nevertheless, I shall suggest that once we have the developed account of Transparency in view, we shall indeed have an explanation of authority[30].

We have, then, a possible motivation for exploring a Transparency account (that is, an account motivated by Evans' remark), namely that on the face of it, transparency seems to explain the apparent groundlessness of self-knowledge, an explanation of which is a desideratum of an account of self-knowledge, and further, an explanation of the authority of self-knowledge is, I suggest, within reach of a properly articulated Transparency account.

A final motivation for a Transparency account is the compatibility of such an account with a (weak) naturalistic picture of the world (and by extension the mind). Nothing in Evans' basic articulation above appeals to any special class of non-natural fact or mechanism to explain the subject's knowledge of her belief. Indeed, as I will suggest, the central insight of Evans' remark is that no mechanism beyond that which explains a subject's knowledge of how things are is needed to explain a subject's self-knowledge of her belief.

[29] This is, of course, barely even a sketch of an account of the apparent groundlessness of transparent self-knowledge. But the aim at this stage is not to provide an account, rather it is to motivate the exploration of the possibility of such an account, and the bare sketch provided suggests that such an exploration is worthwhile. We will see in Chapter three that the Simple Account gives a tidy explanation of groundlessness.

[30] In Chapter three, I will suggest how the Simple Account does indeed provide for authoritative self-knowledge.

2.2. Constraints on a Transparency Account

Underlying Evans' insight is the thought that in forming our beliefs, we are engaged with and responsive to how things are with the world.

But this thought requires further explanation and poses a further question. That I, the writer of this thesis, am engaged with and responsive to the world, and that at least some of my beliefs are therefore responsive to worldly facts does not answer why my *knowledge of myself* would be responsive to worldly facts. We might think that it does not follow from the fact that a subject's beliefs about how things are with the world are sensitive to worldly facts that her beliefs about herself should be sensitive to the same worldly facts. It seems plausible that we can draw a connection between at least some of a subject's beliefs about how things are with the world and the fact that the subject is a person engaged with and responsive to how things are in the world, but it is less clear why there should be a connection between the latter and a subject's beliefs about how things are with herself. This seems to bear a close relation to the Puzzle of Transparency discussed in section 1.4.1: how can the procedure a subject exercises in answering a question about the world possibly answer a question about her beliefs? We will return to the Puzzle of Transparency in section 3.2, but for now we will aim to understand how Evans develops the Transparency Remark.

To understand Evans' development of the Transparency Remark, we should begin by considering the practice a subject *S* engages in when she asserts (as she would put it) 'I believe that *p*'. In asserting 'I believe that *p*', it seems that the asserting subject *S* ascribes a particular property, 'believing that *p*', to a particular object, namely herself, the object referred to by the term in the subject position of the sentence. In other words, the subject engages in *self-ascription*. This is Evans's view. For Evans, the form of belief assertion is the predication of a property to an object. Further, the latter is not only true of the case of a speaker asserting their own beliefs; if a speaker asserts that 'John believes that *p*', she likewise ascribes a property (believing) to an object (John). The use of 'believes' across the first and third personal cases appears to be univocal – the concept picked out by 'believe' when a subject asserts 'I believe that *p*' is, it seems, the very same concept of belief picked out by 'believe' in the assertion 'John believes that *p*'.

Evans' goal for his account of how a subject comes to a belief that *p*, whether this is a first order or a higher order belief, is to preserve this univocity and generality of the concept of belief and respect the form of the assertion of belief as the ascription of a property to an object. Part of Evans' aim in respecting this generality is to preserve a generality of the form of the assertion, i.e., the form of the thought, whereby a property is predicated of an object, and the concept which is used in the predication is univocal across first and third person uses. This points us toward what Evans calls *The*

Generality Constraint. The Generality Constraint is a constraint on the structure of thought. As Evans puts it

> "It seems to me that there must be a sense in which thoughts are structured. The thought that John is happy has something in common with the thought that Harry is happy, and the thought that John is happy has something in common with the thought that John is sad." (Evans, 1982, p. 100)

There is something in common between these thoughts, in that the object of which the property is predicated in the thought that John is happy is the same as that in the thought that John is said, and the property that is predicated of the object in the thought that John is happy is the same in the thought that Harry is happy. Evans suggests that the structuring of thought is to be understood in terms of conceptual abilities:

> "I should prefer to explain the sense in which thoughts are structured, not in terms of their being composed of several distinct *elements*, but in terms of their being a complex of the exercise of several distinct conceptual *abilities*." (Evans, 1982, p. 101)

The subject who thinks 'Harry is happy' and thinks 'John is happy' exercises the same conceptual ability, namely possession and application of the concept *happy*. And the same goes for the exercise of the ability to possess and use the concept *John* in the case of 'John is happy' and 'John is sad'. In this way, in the thought 'I believe that *p*', the thinker exercises the ability to possess and use at least the concept *I*, and the concept *belief*. If we say a subject understands a sentence or has structured thought in terms of the exercise of abilities, Evans suggests that

> "… we commit ourselves to certain predictions as to which other sentences the subject will be able to understand; furthermore, we commit ourselves to there being a common, though partial, explanation of his understanding of several different sentences." (Evans, 1982, p. 101)

We commit ourselves to the explanation that when a subject exercises a conceptual ability, she understands the concept used in such a way that the same concept could be exercised in different contexts or employed by different subjects. In committing to this, we commit ourselves to the claim that the use of a concept has, across all subjects, a common causal-explanatory basis. We must say something about the nature of this causal-explanatory basis in understanding the structure of thought.

The ability to exercise a concept is general, and the structure of thought is understood in terms of conceptual abilities, so it follows from this that there is a generality to thought, which acts as a constraint on the structure of thought. Evans gives the Generality Constraint as follows:

> "[I]f a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception." (Evans, 1982, p. 104)

Evans aims to fulfil two desiderata (which themselves suggest a general constraint on the structure of thought) in giving an account of a subject's belief that *p*. First, that assertions of belief respect a particular form, whereby an object is ascribed with a property, and second that the use of 'belief' is univocal across first and third person uses. A proper understanding of exactly what Evans takes the Transparency Remark to amount to will suggest that his account of the transparency of self-knowledge does indeed fulfil both of these desiderata, and by extension the Generality Constraint[31]. The common causal-explanatory basis for what Evans would call 'concepts of the objective' takes it that to possess such concepts, the subject *S* must be one of the manifold objects in the world, otherwise we could not understand the commonality of the basis of a concept used by the subject *S* and that concept used by another. A consequence of this is that in claiming that 'I believe that *p*', the subject exercises a complex of conceptual capacities to ascribe the property of believing to an object. We must now turn to Evans' understanding of the manner of that ascription.

Consider the ascription of a property to an object (we will continue with the example of belief): 'John believes that *p*.' In asserting this, the subject identifies an object and exercises competence with a particular concept (John), then ascribes a property (belief) to that object, again exercising a conceptual competence with the concept of belief, as suggested above. The same, it seems, is true in the case where the concept exercised in the identification of the object of the ascription is the concept represented by the first-person pronoun, i.e. 'I believe that *p*.' The speaker identifies an object via the first-person pronoun; the very object that they themselves are, through the use of a particular concept (the 'I'-concept) and ascribes this object (the very object they are) with a property (again belief). When the speaker makes such a first personal assertion, they demonstrate their competence with the 'I'-concept and engage in *self*-identification. The speaker understands that in using the 'I'-concept they pick out the very thing they themselves are from the manifold of objects, and they understand that in using such an 'I'-concept, the very same concept could have been used by an arbitrary subject *S* to likewise pick themselves out from the manifold.

So, on this account, the very prospect of a subject being in a position to assert that (as she would put it) 'I believe that *p*' requires that the subject be able to pick out which one she herself is from the manifold of objects via the use of the 'I'-concept. If the assertion 'I believe that *p*' is to be understood as the predication of a property to an object (the object the speaker picks out with the 'I'-concept), the one asserting must understand which object it is that the property is being

[31] I will, however, suggest in chapter three that by fulfilling the first of these criteria, Evans' account is undermined.

predicated of. A speaker's assertion of their own belief is bound up with a self-identification, which is *inter alia* an awareness of themselves as the bearer of the very belief asserted.

2.2.1. Self-Identification

Self-identification is tightly bound up with Evans' account of self-knowledge:

> "We clearly do have ways of gaining knowledge of ourselves, and 'I'-thoughts are thoughts which are controlled, or are disposed to be controlled, by information gained in those ways." (Evans, 1982, p. 207)

In suggesting that in manifesting an 'I'-thought, one manifests an awareness, there is a further suggestion that to manifest an 'I'-thought is to engage in self-reference. 'I'-thoughts are thoughts in which the subject thinks about herself, and the 'I' of the 'I'-thought picks out the subject as the subject of the 'I'-thought. 'I'-thinking is self-thinking and is as such self-referential. The self-referential nature of 'I'-thoughts lends a self-referential nature to self-knowledge:

> "I do not merely have knowledge of myself, as I might have knowledge of a place: I have knowledge of myself *as* someone who has knowledge and who makes judgements, including those judgements I make about myself." (Evans, 1982, p. 207)

For there to be the very possibility that the subject could have self-knowledge, the subject must understand herself in a very special way. 'I'-thoughts and self-identification are bound up with the subject *qua* subject. And not just this – also the subject *qua* object. It is here that we see Evans' debt to Strawson, and in turn, to Kant.

2.2.2. Self-Identification in Strawson

Strawson develops Kant's attack on the Cartesian conception of the soul, the conception that

> "…each of us, by the mere fact of conscious experience, knows that he exists as a Cartesian thinking substance, i.e. as an immaterial, persisting, non-composite, individual subject of thoughts and experiences, capable of existence in total independence of body or matter." (Strawson, 1966, p. 101)

This conception of the Soul (the self) as capable of existence independent of the body is diametrically opposed to the conception of the self that Evans (and Strawson) hold.

Strawson develops Kant's attack on the Cartesian conception by considering the mistake that the conception makes. We are reminded by Strawson that the Kantian conception of experience demands the *transcendental unity of apperception*. The very possibility of the sort of self-conscious awareness 'I'-thoughts require itself requires the transcendental unity of consciousness. 'I'-thoughts require that the elements of the thinker's consciousness are held together into one consciousness.

This unity is what allows it that if I have a pain (a headache say), and I believe I have a headache, the belief and the headache are unified in one consciousness, the consciousness that both has a

headache and believes it has a headache, *my* consciousness. When I have some (occurrent) belief that *p*, that belief is likewise part of a unity of consciousness amongst other (occurrent) beliefs. The requirement that 'I'-thoughts exhibit such a unity is a requirement for the very possibility of the thoughts being 'I'-thoughts at all. The unity of consciousness is *transcendental*: it is a condition of the possibility of a self-conscious subject. This transcendental unity requires "…that a temporally extended series of experiences should have a certain character of connectedness and unity, secured to it by the concepts of the objective." (Strawson, 1966, p.102)

But this transcendental unity is not enough, the unity of 'I'-thoughts secures only a 'formal' notion of the 'I think'. That is, the 'I think' is, as it were, a mere placeholder, it does not mark what we might call a 'genuine' thought. Rather it marks the possibility of thought in general. Transcendental self-consciousness does, however provide "…as it were, the basic ground for the possibility of an empirical use for the concept of the subject of such an autobiography, the concept of the self." (Strawson, 1966, p. 102) Although the transcendental unity of consciousness is sufficient to secure only the formal 'I think', it provides the grounds for the extension into an 'I-concept' which is more than merely formal, where the 'I think' is not merely a mark of consciousness in general.

The transcendental unity of consciousness suggests that experiences have some character of connectedness. This unity of experience holds together a series of experiences which are linked across time. Further, the connections in which the unity consists have a certain *character*, a certain way they are for that which experiences them. Finally, these experiences are secured by "…concepts of the objective…" (Strawson, 1966, p.102). That is, the concepts of the experiences which are unified are *objective*. But we should be clear on what 'objective' means here. In chapter two of part two of *The Bounds of Sense* Strawson draws from Kant's discussion the question of what must minimally be in place in a sufficiently austere conception of experience "…solely in virtue of the fact that the particular items of which we become aware must fall under (be brought under) general concepts" (Strawson, 1966, p. 40).

Strawson suggests in explicating Kant's description of what is left to work with (""the form of the thought of an object in general""(Strawson, 1966, p. 40)) that the notion of 'object' does a lot here. It suggests 'objectivity' in thought, and the idea of objectivity is what we seek to clarify. Strawson suggests that if the concepts of the experiences are objective then there is a difference between 'seems right' and 'is right':

> "To know something about an object, e.g. that it falls under such-and-such a general concept, is to know something that holds irrespective of the occurrence of any particular state of consciousness, irrespective of the occurrence of any particular experience of awareness of the object as falling under the general concept in question. Judgements about

> objects, if valid, are objectively valid, valid independently of the occurrence of the particular state of awareness, of the particular experience, which issues in the judgement." (Strawson, 1966, p. 40)

Judgements about objects are 'objectively valid'. Just because a judgement seems right does not mean that it is right. But this concerns *judgements* where before we were talking about *concepts*. For present purposes, the mark of a concept of the objective is that they obey the Generality Constraint.

Why, though, should this be the case at all? Why should the unity of experience be such that it must be secured by concepts which obey the Generality Constraint? It must be such because the Generality Constraint not only constrains the structure of thought, it ensures that the concepts used by the subject in thought are *contentful*[32]. Above we described the Generality Constraint in terms of ensuring that the use of a concept is univocal. The concept in question (in the example above, 'belief'), when used by different subjects picks out the same thing. When a subject *S* uses the concept 'belief', another subject *R* understands the use of the concept, as the use is univocal – it means the same thing on the lips of both (indeed all) subjects. To understand why this must be so we should understand the Kantian background against which Strawson's, and Evans' insights are to be understood. The Generality Constraint is a constraint on the structure of thought, and there is a Kantian assumption in the background which restricts the *content* of thought. The restriction is this: "Thoughts without content are empty." (Kant, 1929, A51/B75) McDowell, in *Mind and World* elucidates this point clearly:

> "For a thought to be empty would be for there to be nothing that one thinks when one thinks it; that is, for it to lack what I am calling "representational content"." (McDowell, 1996, pp.3-4)

Thought, in general, represents how things are. An empty thought does not represent anything, it is not about anything. As we have suggested above that thought is conceptually structured, we can understand such structuring as our concepts representing, in thought, how things are. So, such empty thoughts would not only be devoid of content, they would be devoid of any structure. It is difficult to conceive of there being such unstructured, empty thought at all. And this, suggests McDowell, is Kant's point. To lack such content and structure "…would be for it to not really be a thought at all, and that is surely Kant's point; he is not, absurdly, drawing our attention to a special kind of thoughts, the empty ones." (McDowell, 1996, p. 4). The unity of experience must be secured by concepts which obey the Generality Constraint (concepts of the objective) because if they did not

[32] This, I think, is also reflected in the seems/is right distinction. If 'seems right' is indistinguishable from 'is right' it is difficult to understand what the content of any particular concept across time would amount to – simply employing the concept would always be correct.

obey the constraint, there would be no thought, i.e., no experience to unify. Kant's point (via McDowell) is that there are no unstructured, empty thoughts.

We discussed above the idea that there is a character of connectedness and unity over time to the transcendental unity of consciousness. The character of connectedness and unity over time provides "…the basic ground for the possibility of an empirical use of the concept of the subject of such an autobiography, the concept of the self." (Strawson, 1966, p.102) To have 'I'-thoughts, the one using the pronoun must have a unified experience of the world, provided by the character of connectedness of their experiences. They must have a temporal autobiography which they can understand as their own, for example by saying 'I believed that *p*, but now I believe that *q*.'

The self-concept is grounded in an autobiographical unity of experience, which is in turn a unity of consciousness. So, the possibility of empirical self-consciousness, and with it the ability to have thoughts which are self-referential, requires as a ground, a unity of experience.

The character of such a unity, of the connectedness of experience, is significant to our enquiry. In particular, what is provided for (and what is not provided for) by the connection of inner experiences. Strawson suggests that the persistent subject of experience through time requires "…empirically applicable criteria of identity…"[33] (Strawson, 1966, p.102), and that such criteria are not provided merely by "…the kind of connectedness of inner experiences provided for by the necessary unity of apperception." (Strawson, 1966, p.102).

The connection of inner experience is, according to Strawson, insufficient to secure empirically applicable criteria of identity. But why might this be so? When one ascribes inner experience, no criteria of identity is invoked at all. When one uses the first-person pronoun 'I', one does not need to appeal to any criteria of identity to justify one's use of the pronoun. As Strawson (rightly) puts it "It would make no sense to think or say: *This* inner experience is occurring, but is it occurring to *me*?" (Strawson, 1966, p.103) There is, within the realm of inner experience, simply no question of whether the experience belongs to oneself or to another, and as such no criteria of identity need be invoked at all. The question 'to whom is this inner experience happening?' simply never occurs (and indeed it might be incoherent to even ask such a question). Inner experience, when had by a subject, can be flawlessly ascribed to that subject by that subject. The very possibility of inner experience

[33] Strawson takes it that the question as to whether there are such criteria of identity at all is a settled one –

> "…our ordinary concept of personal identity does carry with it empirically applicable criteria for the numerical identity through time of a subject of experiences (a man or a human being) and that these criteria, though not the same as those for bodily identity, involve an essential reference to the human body." (Strawson, 1966, p.102)

At this stage we will not quibble with the settledness, or otherwise of this question

entails an experiencing subject, and this entailment guarantees that when that subject asserts their inner experience, they do so without appealing to any empirical criteria of identity such that they could ask the question 'who is experiencing?'.

This would seem to undermine the requirement that there be such criteria for the possibility of empirical self-consciousness. If one can engage in reflection on inner experience without appeal to empirical criteria, surely one is engaging in self-conscious thought *without* the very empirical criteria of identity Strawson claims are necessary to the possibility of 'I'-thoughts. The key insight, suggests Strawson, is that such criterionless ascription is an illusion. The root of this insight is the thought that even when a speaker uses 'I' without the possibility of such a use being justified by empirical criteria, it *does not thereby lose its role as a referring term*[34] – 'I', when used by the speaker still refers to that very speaker. Strawson suggests that this is the case (perhaps) because it is uttered "…publicly from the mouth of a man who is recognizable and identifiable as the person he is by the application of empirical criteria of personal identity." (Strawson, 1966, p.1 03) The utterances, the uses of 'I' as the marker of the possession of the 'I'-concept are public. The criteria that settle it to whom the use of 'I' belongs, the one who possesses the 'I'-concept, are likewise public, empirically available criteria. When a subject *S* asserts 'I believe that *p*', the criteria that settle it that the 'I' in 'I believe that *p*' refers to *S* are empirical criteria, available to another subject *R*. Further, *S*, if pressed, would

> "…acknowledge the applicability of those [empirical] criteria in settling questions as to whether he, the very man who now ascribes to himself this experience, was or was not the person who, say, performed such-and-such an action in the past." (Strawson, 1966, p.103) [clarification mine].

The criteria are those criteria that settle it that when *S* uses the 'I'-concept, it picks out the same self across time.

The *ground* of the possibility of 'I'-thoughts, for Strawson, still requires that the subject *S* understand both that there are such empirical criteria of identity, and that in order to use the first-person pronoun at all, such criteria apply to her, the subject *S*.

[34] Strawson's understanding of what it is for the first-person pronoun to refer seems to be that it obeys the truth-conditional reference rule: ''I am F' is true just in case the speaker is F'. We shall see in chapter three that one can agree that uses of the first-person pronoun can obey the truth-conditional reference rule but is not (as Anscombe puts it) an expression whose logical role is to make a reference, at all. (Anscombe, 1981, p. 32)

In the case of criteria-free identification, Strawson suggests that we are afflicted by a philosophical illusion:

> "It is easy to become intensely aware of the immediate character, of the purely inner basis, of such self-ascription while both retaining the sense of ascription to a subject and forgetting that immediate reports of experience have this character of ascriptions to a subject only because of the links I have mentioned with ordinary criteria of personal identity." (Strawson, 1966, p.103)

The illusion is this: there is a use of 'I' which is unique to inner experience but is still subject-referring. It can seem that such a use exists (indeed this would be the paradigm use of 'I', exemplified by e.g., the Cogito), but this is an illusion. Were such a use to exist, "…what we really do is simply to deprive our use of "I" of any referential force whatever." (Strawson, 1966, p.103) Were such a use to exist, the use of 'I' would be purely a formal one – there are no criteria under which a subject could be identified, the use of 'I' would be groundless. This purely formal use of 'I' would not pick out a particular subject with a particular autobiography, rather it would "…express, as Kant would say, "consciousness in general"." (Strawson, 1966, p.103) 'Consciousness in general' here suggests that the use of 'I' (were the use to be purely formal) would merely be a marker that the user is a thinker, but not a marker that it is *any specific thinker*; the identification of a specific thinker, the possessor of a particular autobiography, who exists in the spatiotemporal order, requires empirical criteria of identification. For there to be a use of 'I' which is more than this merely formal use, there then must be criteria of identity, which are "…supplied by our ordinary concept of a person as something which, *inter alia*, is an object of outer sense." (Strawson, 1966, p.104) And, significantly for our project, this non-formal use is the use which, for Evans and Strawson, underpins (substantive) self-knowledge. Such a use would conform to the Generality Constraint. When one makes a claim as to whether one believes that *p*, one does so in such a way that the claim to belief is a substantive claim. If the 'I'-thought which underpins the 'I believe that *p*' (where *p* is some empirical or world directed claim) is merely the formal use of 'I', there is nothing which secures it that any specific empirical self believes that *p*. If there is no way to secure it that any specific empirical self believes that *p*, then there is nothing which secures it that the holder of the belief formed the belief with appropriate connections to how things are.

2.2.3. Self-Identification and Transparency

With the Strawsonian/Kantian background in place, we have the demand that in order to think 'I'-thoughts, certain criteria regarding the structure and demands of self-identification must be met. The demand that empirical criteria of identity exist in order that a thinker might have the 'I'-concept, and that the use of concepts (and the structure of thought) obey something akin to the Generality

Constraint moves us toward an understanding of what might motivate Evans' development of his Transparency account of belief.

2.3. Evans' Account

Having established the Kantian background of Evans' remark, we can now examine how he aims to develop it. Evans' own account builds from his account of self-identification and is strongly influenced by the Strawsonian/Kantian concerns elucidated above.

Evans suggests that our idea of ourselves (our 'I'-Idea) consists in a link between certain thoughts ('I'-thoughts) and certain information, information gained through particular first personal channels. But this is not all self-identification requires. It also requires knowledge of the truth of the following identity:

> "…⌜ I= δ_I ⌝. […] where δ_I is a fundamental identification of a person: an identification of a person which – unlike one's 'I'- identification – is of a kind which could be available to someone else." (Evans, 1982, p. 209)

It is here that we see the demand for empirical criteria of identity in forming 'I'-thoughts. In order to have an understanding of what it is to self-attribute some predicate ('…is dead', say), one must couple one's "…general understanding of what it is for a person to satisfy the predicate 'ζ is dead'" (Evans, 1982, p. 209) or the general formulation 'ζ is F', and one's Idea of oneself (one's 'I'-Idea[35]). In order to think of oneself as fulfilling a particular predicate, one must understand both what it is for ⌜ δ is F ⌝ to be true and understand what it is for ⌜ I= δ ⌝ to be true. The understanding of what it is for ⌜ δ is F ⌝ to be true seems at least somewhat uncontroversial – it is to understand what it would be for some arbitrary person in the world, a member of the spatio-temporal order, to fulfil the predicate F (to exhibit F-ness). In the case of the truth of the identity ⌜ I= δ ⌝, it is less clear that this is uncontroversially true, as we will see in the later discussion of Anscombe in section 3.3. For now, however, we have an account of self-identification in view and can turn to the transparency of belief.

2.3.1. Transparency of Belief

We can now begin to understand Evans' notion of mental self-ascription. When one answers the question 'is it the case that *p*?' positively or negatively, one judges that *p* (or not-*p*). Evans suggests that to understand the judgment 'I believe that *p*' one must "…possess a psychological concept expressed by 'ζ believes that *p*', which the subject must conceive as capable of being instantiated otherwise than by himself." (Evans, 1982, p. 226) This determines what sort of evidence one allows to bear on the belief that A believes that *p*, for an arbitrary subject 'A'. The evidence must be such

[35] I take Evans' notion of the 'I'-Idea to be akin to one's self-concept. The idea one has of oneself as oneself.

that it could bear on the deliberation of an arbitrary subject in the spatio-temporal order. For the 'I'-thought which accompanies the belief that *p* to be genuinely contentful – to relate to a specific thinker in the world, there must be, as suggested above, empirical criteria of identity in place to secure which thinker is doing the thinking. Further to this, in making sense of Strawson's insight, Evans suggests that were one to merely look inward to understand whether one believes *p*, it seems that such an inward glance would not properly locate oneself as a thinking subject in the world, precisely because such evidence would emphatically *not* bear on the deliberation of an arbitrary subject. Such evidence would be completely bound up with one's own 'I'-Idea and as such would not be available to an arbitrary subject. We have a further argument from Evans as to the impossibility of the inward glance by examining perceptual beliefs.

The case of self-ascription of perceptual beliefs ('I see an F'), while not quite of the same form is still explained via attending to the world, rather than looking inward. To see why we need to understand how Evans characterises perceptual knowledge. Evans characterises perceptual knowledge as

> "…an information state of a subject: it has a certain *content* – the world is represented a certain way and hence it permits of a non-derivative classification as *true* or *false*." (Evans, 1982, p. 226).

Further, the judgements in which perceptual knowledge consist are "…*based upon* (reliably caused by) these internal states." (Evans, 1982, p. 227) So the picture is of one moving from an experience (an informational state) to a judgement. One important thing to note is that the subject bases their judgements on their experience, they do not make judgements about their experience (the informational state). Rather the process of judgement (or conceptualisation) is a movement from one kind of cognitive state (with one sort of content – non-conceptual content which the informational state consists in) to another (with conceptual content). Then, when the perceiving subject wishes to check their judgement, they gaze again at the world to reproduce the informational state which was conceptualised in their act of judging that *p*. They look outwards, not inwards. Indeed, a subject could not 'look inwards' to check his state – "[h]is internal state cannot in any sense become an *object* to him. (He is in it)." (Evans, 1982, p. 227)

To gain knowledge of internal states the subject uses the same skills of conceptualisation they would in gaining knowledge of the external world. The subject goes

> "…through exactly the same procedure as he would go through if he were trying to make a judgement about how it is at this place now, but excluding any knowledge he has of an *extraneous kind*. […] The result will necessarily be closely correlated with the content of the informational state which he is in at that time." (Evans, 1982, p. 227-228)

The correlation between the result of the judgement of how things are here and now and the content of the informational state the subject is in lets the subject produce (and give expression to vocally through locutions such as 'It is as if to me that') cognitive states systematically dependent on the content of information states, which is a basis for a knowledge claim regarding the informational state. But as with the perceptual case, the state is still not an object to the subject –

> "…there is nothing that constitutes 'perceiving that state'. What this means is that there is no *informational* state which stands to the internal state as that internal state stands to the state of the world." (Evans, 1982, p. 228)

That is, there is no state which is such that even if the subject were to look inward, the subject could use this state as a basis to gain knowledge of their internal states. The subject, in order to gain knowledge of their internal states, must look outward. Thus, we can see the motivation for (and something of the explanation of) the Transparency Remark. Were the subject to 'look inwards' to find out what her belief is, there would be no informational state to be the basis of her assertion that (as she would put it) 'I believe *p*'. To return to the original formulation of the Transparency Remark, the subject answers the question 'do you believe that *p*?' not by looking inward (and by Evans' lights, finding nothing), but by looking outward and answering the question 'is it the case that *p*?'

We can shift this talk of informational states and the transition of states that Evans considers judgement to consist in to talk of the epistemic basis of assertion. Evans' idea here seems to be that the 'informational state' which is correlated with the content of the assertion in question can be understood as the epistemic basis of the assertion. Evans' point that if the subject were to look inwards there is no state which could be a basis for the assertion in question. This is not merely the point that self-knowledge is immediate or groundless (i.e., based on nothing). Rather, Evans' point suggests that self-knowledge *cannot* be groundless or immediate if immediacy is understood in these terms. The point is not that there is no epistemic basis for assertions of the form 'I believe *p*'. Rather the point is that the epistemic basis of the assertion 'I believe that *p*' can be nothing more than the epistemic basis of the assertion '*p*', and the epistemic basis of the assertion '*p*' is not made available by exercising an epistemic capacity whose exercise consists in an introspective 'inward glance', but by exercising the very same epistemic capacity exercised in looking outwards at how things are with the world. The assertion 'I believe that *p*' is groundless or immediate insofar as the epistemic capacities exercised in making the assertion are nothing *more* than the epistemic capacities exercised in making the assertion '*p*'. *This* is Evans' key insight. The talk of 'informational states' and 'internal states' serve to obscure this point, but when we re-engineer Evans' thought in

the terms of epistemic capacities, a central thread emerges, consistent with the Kantian (and anti-Cartesian) background of the Transparency Remark.

2.4. Concluding Remarks to Chapter 2

In this chapter, I have aimed to explicate not only Evans' development of the Transparency Remark via his account of self-identification, but the theoretical background upon which the development of the Transparency Remark in Evans rests. In this background, we can see, I suggest, the scope for a simple, relatively austere account of Transparency which takes Evans' central insight that the epistemic capacities exercised in making the assertion 'I believe that *p*' are nothing more than the epistemic capacities exercised in making the assertion '*p*'. What I have not done in this chapter is engage with the Puzzle of Transparency (or substantively with any of the objections to a Transparency account, but I take the Puzzle to be the central objection). Indeed, we still do not have a fully satisfactory characterisation of the Puzzle. In chapter three I will engage with the Puzzle of Transparency in depth, and in doing so, develop what I call the Simple Account of Transparency — an account of Evans' remark which develops the central insight revealed in this chapter and answers the central objections to Transparency. In answering the Puzzle of Transparency, however, we will need to reject Evans' account of the 'I-idea' and self-ascription, which will present a challenge: we must still provide for an appropriate generality of thought.

3. The Simple Account of Transparency

I discussed Evans' own understanding of the Transparency Remark in the previous chapter. The aim of this chapter is to develop what I take to be the central point of Evans' remark — that the epistemic capacities exercised in making the assertion 'I believe that *p*' are no more than the epistemic capacities exercised in making the assertion '*p*' — into what I call the Simple Account of Transparency. In developing this account, I will engage with what I take to be a clarified version of The Puzzle of Transparency discussed in section 3.2. Answering this puzzle will expose the central difference between the Simple Account and Evans' own account of the Transparency Remark. I will also discuss how the Simple Account might fend off the other core objections to a Transparency account.

3.1. Epistemic Procedures and Epistemic Capacities

Recall that Evans suggests:

> "If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the question whether p." (Evans, 1982, p. 225)

Evans' remark suggests a simple thought; we gain knowledge of at least some of our mental states, not by introspection, by 'looking inward' at ourselves, but rather by 'looking out' at the world. That is, to answer a question whose topic is how things are with myself (namely whether I believe *p*), I put into place the procedure I would engage in to answer a question whose topic is how things are with the world (namely whether it is the case that *p*).

One way to understand Evans' insight is by taking seriously his notion of the same procedure being used to answer both questions. Evans' suggestion is that the very same procedure used to answer the question 'Is it the case that *p*?' is used to answer the question 'Do you believe that *p*?' My looking out of my window gets me into a position to answer not only the question 'is it raining' but also the question 'do you believe it is raining?'. Evans' suggestion is that there is a single epistemic procedure that can provide an answer to two questions with distinct topics: the topic of how things are with the world and the topic of how things are with the subject.

It may seem that we should begin the explication of Evans' remark by thinking hard about what the epistemic procedure in question might be. This, I think, would be a mistake; Evans' position is (broadly) agnostic on what the procedure itself amounts to. What is important is that the same procedure delivers answers to both questions, not what that procedure amounts to. Throughout this chapter I will talk of both epistemic procedures (following Evans' formulation of the Transparency

Remark) and epistemic capacities. I will use these terms interchangeably (except when I specify otherwise). By 'epistemic procedure' and 'epistemic capacity' I mean something the actuality of which affords an answer to a question, e.g., to a question of the form 'Is it the case that p?' The difference can be encapsulated in the following: an epistemic procedure is what the subject engages in when she exercises an epistemic capacity[36].

3.1.1. Questions, Answers and Assertions

It is helpful to clarify things further here by thinking not of the questions 'is it the case that *p*?' and 'do you believe that *p*?', but rather by thinking of the assertions that constitute the positive answers to those questions. That is, rather than thinking about the same epistemic procedure being used to answer both the question 'is it the case that *p*?' and the question 'do you believe that *p*?' we should think about the same epistemic procedure or capacity being exercised in the making of the assertions '*p*' and 'I believe that *p*'[37] Indeed, Evans is clear that the answers to are to be understood as assertions:

> "We can encapsulate this procedure for answering questions about what one believes in the following simple rule: whenever you are in a position to assert that p, you are *ipso facto* in a position to assert 'I believe that *p*'" (Evans, 1982, pp. 225-6)

This is worth being clear about as the discussion in relation to the Puzzle of Transparency in section 3.2. will shift to the link between the two assertions which constitute answers to the questions, so clearing the ground now helps to focus on that later.

Boyle's (2011) discussion furnishes us with one other important idea — that when a speaker asserts '*p*' she need do nothing more than reflect to pass to a position where she can assert 'I believe that *p*'. Of course, 'reflection' is still somewhat obscure. Evans' remark that Transparent self-knowledge means that when one is in a position to assert '*p*' one is *ipso facto* in a position to assert 'I believe that *p*' can help clarify things for us. We might think of Boyle's notion of needing to do nothing more

---

[36] Much of the inspiration for this account is drawn from Boyle (2011). There Boyle suggests that the Transparent Procedure be understood as the actualization of a cognitive power:

> "[T]he important truth is this: the very same actualization of my cognitive powers that is my believing P is, under another aspect, my tacitly knowing that I believe P. Hence, to pass from believing P to judging I believe P, all I need to do is reflect – i.e., attend to and articulate what I already know. Something broadly similar will hold for other psychological conditions of which I can have transparent self-knowledge." (p. 6)

However, Boyle's formulation contains the obscuring notion of believing *p* being tacitly knowing *p* under another aspect. We do not need this notion to understand Evans' remarks and they only serve to obscure the point. Nevertheless, Boyle's discussion of a 'cognitive power' aims at the same ideas I develop in terms of epistemic capacities.

[37] There is a flat-footed response that the positive answer to both of these questions is 'yes'. This is too quick. The 'yes' can be understood as 'yes, *I believe that p*' or 'yes, *p*'. The 'yes' is already tacitly included in the assertions above if they are understood as answers to questions, and 'yes' as an answer to the questions tacitly includes the appropriate assertion.

but reflect on one's assertion and Evans' idea that one is in virtue of asserting '*p*' *ipso facto* in a position to assert 'I believe that *p*' as an *entitlement* to move from one to the other, and an entitlement one has in virtue of doing nothing more than making a certain sort of assertion. Boyle introduces the idea of reflecting on the assertion, which suggests one performs an epistemic procedure or exercises an epistemic capacity beyond merely asserting '*p*' to move to asserting 'I believe *p*', but this, I think, is the result of a loose formulation. The idea of 'entitlement' and 'movement from one answer to the other' also suggests the subject engages in a further epistemic procedure, exercises a further epistemic capacity. Thinking this would be a mistake. The reflection in question, the entitlement to answer the belief question by doing nothing more than answering the question of how things are is not an engagement of a further epistemic capacity to know what one is up to. There is a sense in which the talk of 'reflection' and 'entitlement' are in a sense obscuring of the simple central point that Evans is making. We should throughout think of them as heuristics or loose formulations of the simple idea that the answer to the question 'do you believe that *p*?' is made by exercise of no further epistemic capacities than those exercised in answering the question 'is it the case that *p*?'

3.2. A Puzzle of Transparency

We are, however, presented with a puzzle: how can an answer to a question regarding how things are with the world possibly lead to an answer to the question regarding how things are with the subject? How can the epistemic procedure engaged in answering the question 'is it the case that *p*?' possibly put a subject in a position to answer the question 'do you believe that *p*?'? This is the Puzzle of Transparency, and any satisfactory explication of Evans' insight must answer or dissolve this puzzle. But to treat of the puzzle satisfactorily, we must formulate it more sharply. The puzzle, loosely formulated, 'how can one procedure deliver answers to questions on two distinct topics', or 'how can the epistemic procedure exercised in the assertion "*p*" be the same as the epistemic procedure exercised in the assertion "I believe that *p*" given that one assertion concerns how things are with the world and the other concerns how things are with the subject'? Neither of these formulations will do, although the second formulation is closer to a useful one.

3.2.1. Sharpening the Formulation – Semantic Discontinuity

A sharper formulation should understand what we mean by the questions having distinct topics or the answers having different concerns. The puzzle asks how exercising one epistemic procedure can provide an answer to two questions on distinct topics. To sharpen the puzzle, what we want to understand is what we mean by the topic of an assertion (i.e., the answer to a question), or the concern of that answer. We might think that the topic of an assertion is given by the content of that assertion. The topic of my assertion 'it is raining' is, one would think, the rain, and the content of the

assertion concerns the rain, whereas the topic of my assertion 'I believe it is raining' is my belief, and the content concerns my beliefs. So, what is needed in a sharper formulation is a formulation which is sensitive to the individuation of contents. Framing the discussion in terms of the *truth conditions* of the assertion that constitutes the answer to the question gives us this sharper formulation. We can, it seems, demarcate assertions by truth conditions and capture the notion of 'topic' or 'concern' in the loose formulation of the puzzle. So, The Puzzle of Transparency becomes the puzzle of how two assertions made by exercise of the same epistemic procedure or capacity can have different contents, that is different *truth conditions*. After all, the assertion '*p*' is true just in case *p* and the assertion 'I believe *p*' is true just in case the speaker believes *p*. The Puzzle of Transparency is the puzzle of how it can be possible for these truth-conditions to be distinct if the assertions are made by putting in to place the same procedure. After all, we would think that if the assertions are made by the exercise of one epistemic procedure, then they would be made true in the same circumstances. If the procedure by which the assertion 'I believe that *p*' is made is the same as the procedure by which the assertion '*p*' is made, what settles it that the 'I believe…' in 'I believe that *p*' makes a contribution toward the content of the assertion, such that the truth-conditions of the assertions 'I believe that *p*' and '*p*' can be distinct? The Puzzle of Transparency asks how it can be that the assertion 'I believe that *p*' is *semantically discontinuous* with the assertion '*p*' given they are asserted by putting in to place the same epistemic procedure. Notice, however, that here we have (quite naturally) aligned the very idea of the content of an assertion with the idea of the truth-conditions. We shall see that there is reason to question this alignment, and that this affords a way of responding to the Puzzle. Our response will come from what seems like a surprising direction – Anscombe's discussion of the first-person term, and in particular what distinguishes the first-person term from a special sort of name, a name that each one of us has only for ourself. We shall see, however, as our inquiry progresses, that Anscombe's discussion figures centrally in a proper understanding of the Transparency Remark.

3.3. Toward an answer: Anscombe and the A-Practice

I want to suggest that we can make headway here by considering G.E.M. Anscombe's infamous remark from *The First Person*:

> "…this is the solution: "I" is neither a name nor another kind of expression whose logical role is to make a reference, at all. Of course we must accept the rule "If X asserts something with 'I' as subject, his assertion will be true if and only if what he asserts is true of X." But if someone thinks that is a sufficient account of "I", we must say "No, it is not", for it does not make any difference between "I" and "A". The truth condition of the whole sentence does not determine the meaning of the items within the sentence." (Anscombe, 1981, pp. 32-33)

Anscombe is here suggesting that what we might call the 'truth-conditional reference rule' for the first-person pronoun[38] is insufficient to capture the distinctiveness of assertions with 'I' in the subject position when compared to assertions with 'A' in the subject position[39]. 'A' here is understood as a special name that each speaker uses only for themselves. The distinction between 'A' and 'I' is this: to make an assertion with 'A' as subject the asserter must exercise the epistemic capacity of identification of an object, but in assertions with 'I' as subject, no such epistemic capacity is exercised. To make sense of Anscombe's point we must take a short detour into the 'A-practice'[40].

The aim of Anscombe's 'A-practice' example is to show that 'I' does not have the same properties as a special name each one uses to refer to themselves, and as such cannot be treated as such a name. She asks us to imagine a people who have no first-person pronoun as we would understand it. Instead, each one has a name, marked on their body in an area not visible to them (between their shoulders and at the top of their chest). This name is the name that others use for them. Additionally, each of them has on the inside of their wrist a marking, 'A', which is the special name each one of them has for themselves. When an 'A'-user asserts 'B is F', they identify the subject B by the name on their back or chest with the demonstrative identifications 'that man is B' and (of the very same object) 'that man is F' and having the two demonstrative identifications 'that man is B' and 'that man is F' of the same object, they infer that the man that is B is the man that is F, or, 'B is F'. The 'A'-user's assertion 'B is F' is grounded in an identity judgement (namely, 'that man is B'), and exercises the epistemic capacity of observational identification (and of inference).

When an A-practice user must assert something regarding themselves, such as in the assertion 'A is F' they likewise look to a marking on the body in order to identify the subject. And because the marking on the chest or back is not visible to the asserting subject, they use the marking on their wrist to find out which name to use and demonstratively identify 'that man is A'. Likewise, they identify through a demonstrative 'that man is F' and infer 'A is F' on the basis of these two demonstrative identifications. The epistemic capacities the 'A'-user engages in the assertion 'B is F' are those associated with using a name, and the suggestion is that the epistemic capacities used in the assertion 'A is F' are, in this respect, the very same capacities. The important difference between the assertion 'A is F' and the assertion 'B is F' is not in the epistemic capacities exercised in making the assertion, but rather that "…for each person there is one person of whom he has characteristically limited and also characteristically privileged views." (Anscombe, 1981, p. 24).

[38] An assertion of 'I am F' is true just in case the speaker is F.
[39] I will throughout use the locution 'assertions with 'I' as subject'. We should understand 'as subject' to be equivalent to 'in the subject position of the assertion'.
[40] I present an abbreviated version of the argument. See Anscombe (1981), pp. 24-27.

When identifying the very person they are, an 'A'-user can only use the name on the wrist, not the chest or back. Nevertheless, they exercise the same epistemic capacities as using a name to refer to another. The assertions 'A is F' and 'B is F', for the 'A' user always exercise the epistemic capacity of observational identification of an object. We should note here that there is something fundamentally *odd* about the A-practice, in that the A-practice users are only capable of observational identification. The point we should take about names from the A-practice is that the use of a name opens the possibility of observational identification of the subject of an assertion. It might be that the fundamental grounding of the use of a name is a demonstrative identification (when the name is first used), but we need not commit to that, merely that the use of 'A', or any other name, opens up the possibility of observational identification. Anscombe's point is that 'I' does not function like this. Making an assertion with 'I' as the subject does not exercise the epistemic capacity of observational identification to settle which thing that subject is. When a speaker uses 'I', there is no observation of the one is being identified; when a speaker uses 'I' from time to time, they do not observationally identify (or re-identify) the one they are.[41]

We can see this by considering that the idea of identification (and re-identification) introduces the possibility of *misidentification*, and it is a widely accepted property of the first-person pronoun that assertions with the first-person pronoun as subject are *immune to error through misidentification*[42][43]. Take an assertion 'I am F'. The intuitive thought is that in asserting 'I am F' I cannot be wrong about the one to whom I attribute F-ness. Contrast this with an assertion with a name in the subject position, like 'John is F'. The assertion with a name in the subject position can go wrong in two different ways: The asserter can mistakenly attribute F-ness to John, for perhaps John is not F but G. The asserter has correctly picked out John but have attributed to him the wrong property. But the asserter can also go wrong in a different way. They can correctly attribute being F to someone, but that someone *is not John*. Perhaps, for example, it is John's twin brother James who is F, and the one making the assertion has mistaken John for James[44]. In this case, the asserter has picked out the incorrect object but has attributed to it the correct property. This error is possible

[41] We might ask how it can be that someone who is ostensibly a subject cannot tell which one they are without checking observationally? This intuition is, I think, grasping at the point that 'I' does not identify an object *at all*.

[42] For detailed treatments of immunity to error through misidentification, see Pryor (1999), Shoemaker (1968), Wittgenstein (1958)

[43] This is not *strictly* true – not all assertions with 'I' as subject are immune to error through misidentification, but all assertions with 'I' as subject are immune to error through misidentification *insofar as they are not based on identity judgements*. This is a subtle wrinkle which can be passed over without further comment as the assertions which form the concern of this work are not based on identity judgements.

[44] The asserter could also assert that 'John is F' when in fact James is G. This would not be a third sort of error but would rather be both an error of predication and an error in getting the right object.

because of the possibility of inferentially grounding assertions with a name in the subject position; like the case of the 'A'-user asserting 'B is F', the assertion 'John is F' can be grounded in the demonstrative identification 'that man is John' and the demonstrative identification of the very same thing 'that man is F', leading to the inference 'John is F'. The assertion with a name as subject can, in principle, go wrong at two places, the identification of someone as John, and the identification of someone as F, and as such the inference can fail. We can see something of an explanation of why the first-person pronoun enjoys immunity to error through misidentification by examining the difference between 'I' and 'A'. We have already suggested that to make an attribution of a property, an 'A'-user must exercise the epistemic capacity to observationally identify via a demonstrative (i) 'that one is A' and (ii) 'that one is F' and (iii) make an inference that the demonstrative identifications in (i) and (ii) entail that 'A is F'. This is contrasted with an assertion with 'I' as subject, for whereas assertions with names as subject provide for the possibility of their being grounded inferentially in the manner just described, assertions with "I" as subject do not. As such, whereas assertions with names as subject may be said to be immune to errors through misidentification -- in that they can go wrong on account of the identity judgment on which they are grounded-- assertions with "I" as subject may be said to be immune to errors of this kind[45].

The point of the 'A'-practice example should now be clear. Anscombe's point is that treating 'I' as a special sort of name introduces the possibility of making an error through misidentification, as using the special name 'A' exercises epistemic capacities not exercised in the use of 'I', capacities the exercise of which entail the possibility of the exercise going wrong, i.e., of misidentifying. As such, 'I' should not be treated as a special sort of name. That the use of 'I' as subject in an assertion exercises no epistemic capacities for observational identification is brought into sharp relief by the following thought: there is no cognitive achievement on the part of the speaker in using 'I' as subject in an assertion. A speaker who asserts 'I believe that *p*' does not achieve anything by picking out the right subject of her assertion. If the asserting subject had exercised an epistemic capacity to observationally identify which one 'I' is, she could have gone wrong – the capacity could have misfired in some way, and she could have misidentified which one 'I' is[46], and as such getting hold of the right thing would be an achievement. This presents us with another puzzle, one which we will put off engaging with for now. The puzzle is this: Surely *NN* could go wrong in her assertion 'I believe that *p*', because she does not believe that *p*. This suggests that her assertion 'I believe that *p*'

[45] There would also be a possibility of reference failure, which we will discuss below.
[46] This short argument is similar in form to Shoemaker's arguments against self-blindness. Briefly, the capacity to know one's own mind cannot be via an introspective equivalent of perception, as perception entails the possibility of systematic error, i.e., blindness. The idea of being 'introspectively blind' is, Shoemaker suggests, incoherent, so introspection cannot be perceptual in this way. See e.g. Shoemaker (1996)

exercises some epistemic capacity over and above her assertion '*p*' because, presumably, she could assert '*p*' and assert falsely 'I believe that *p*'. We will return to this in section 3.4.2.

3.3.1. 'I', 'A' and Epistemic Capacities

'I' and 'A' can be substituted with one another in their respective truth conditional reference rules and the truth of the assertions is unaffected. The respective rules are:

> A: An assertion 'A is F' is true just in case the speaker is F.
>
> I: An assertion 'I am F' is true just in case the speaker is F.

Nevertheless, the discussion so far has cemented the idea that 'A' and 'I' are distinct in a way which is not captured in the truth-conditional reference rule, and that way seems to be related to the epistemic capacities that are exercised in the making of the assertions in question.

Anscombe suggests the distinction between 'I' and 'A' is that of *self-consciousness*:

> "The first thing to note is that our description does not include self-consciousness on the part of the people who use the name "A" as I have described it. They perhaps have no self-consciousness, though each one knows a lot about the object that he (in fact) is; and has a name, the same as everyone else has, which he uses in reports about the object that he (in fact) is." (Anscombe, 1981, pp. 24-25)

It is not clear what 'self-consciousness' amounts to here, but it is clear that it is manifested in assertions with "I" as subject on account of the fact that they exercise no epistemic capacities for observational identification of the subject of the assertion. I have suggested above that assertions with 'A' as subject are inferentially grounded in an observational, demonstrative, identification of an object. This grounding itself goes beyond the truth-conditional reference rule; there is nothing in the rule that demands or mandates anything about an observational identification. Indeed, the application of the truth-conditional reference rule already takes it that an identification has taken place, in that, in order to apply the rule, it must be that the one applying it has already exercised the epistemic capacity of observational identification: Application of the rule takes it that someone has been identified as the speaker ('that is the speaker') and further it ascribes a property to that speaker, the property of having made an assertion ('that one asserted "A is F"' —the assertion *mentioned* in the rule). The application of the truth-conditional reference rule to a specific assertion exercises the same epistemic capacities exercised by the 'A'-user in her assertion 'A is F'. This is unlike the 'I'-user, who does not exercise the epistemic capacity for observational identification in her assertions with 'I' as subject, and as such does not exercise the same capacities in making her first-personal assertions as in her application of the truth rule to those assertions. This may make it

seem like the 'A'-user is doing more, achieving more (epistemically) than the 'I'-user. After all, each use of 'A' as subject in an assertion is an epistemic achievement on the part of the 'A'-user. But this, I think, tells us something significant about 'self-consciousness' as understood by Anscombe: self-consciousness is not a cognitive achievement. Insofar as an 'I'-user has non-observational knowledge of the thing they are, that knowledge is not had through the exercise of an epistemic capacity for observational identification (on pain of error), and as such seems to be no achievement at all. Rather, it is something they have in virtue of being an 'I'-user.

3.3.2. Transparency and Anscombe's Point

The link to the Simple Account of Transparency should now be coming into focus. The assertion of 'I believe that *p*' exercises no epistemic capacities beyond the epistemic capacities involved in the assertion '*p*'. In particular no capacity for observational identification of the subject is involved in the assertion 'I believe that *p*'. But this does nothing to resolve the puzzle of semantic discontinuity above. If anything, it makes the puzzle sharper; the content of the assertion 'I believe that *p*' is surely more than the content of the assertion '*p*', so how can the assertion which contains the 'I believe that…' exercise no more epistemic capacities than the assertion that '*p*'? We can dissolve the puzzle by wholly grasping Anscombe's thought that "'I" is neither a name nor another kind of expression whose logical role is to make a reference, at all." (Anscombe, 1981, p. 32). What Anscombe means by 'reference' here can be understood as identifying knowledge of an object. As Haddock (2019) puts it

> "The knowledge that the user enjoys, in knowing which object this is, is identifying knowledge. It is not merely general knowledge to the effect that the expression in subject position refers to something that belongs to a certain general kind, or satisfies a certain general description. […] To reject the idea that "I" is an expression whose logical role is to make a reference is to reject the idea that a sentence with "I" as subject is a sentence whose uses contain this identifying knowledge." (p. 958)

That is, assertions with 'I' as subject make no identification of the one upon whom, according to the truth conditional reference rule, they turn to for their truth. If it did, the identifying subject would, it seems, enjoy identifying knowledge of that which was identified, and there is no such identifying knowledge, so no identification. So far, we have ruled out that 'I' acts a special sort of name due to assertions with 'I' as subject being immune to error through misidentification, unlike even the very special name 'A', but this does not rule out that assertions with 'I' as subject furnish the 'I'-user with identifying knowledge. We cannot yet rule out that the 'I'-user has identifying knowledge because we might think that 'I' could be subsumed into a special sort of demonstrative, since demonstrative reference is also immune to error through misidentification, and demonstrative reference still seems to employ identification of an object. This will not do; assertions with first-person pronoun as subject

also have the property of being immune to *reference failure*, unlike assertions with demonstratives in the subject position.

Take a case like this: Our speaker, *NN* asserts 'that man is Elvis Presley' while pointing at something. Unbeknownst to *NN*, the thing she demonstrates with 'that man' is not a man at all but is in fact a cleverly arranged cardboard cut-out. *NN*'s assertion does not merely attribute being Elvis Presley to the wrong thing, nor does it mistake one man for another. Rather, her demonstrative 'that man' fails to refer to anything at all[47]. This is because there is a conception of what is indicated by the demonstrative internal to the demonstrative itself. As Anscombe puts it

> "…even though someone may say just "this" or "that", we need to know the answer to the question "this what?" if we are to understand him; and he needs to know the answer if he is to be meaning anything." (Anscombe, 1981, p. 27)

It is internal to the demonstrative that there is an answer to the question 'this what?' – a conception of what is being demonstrated. The demonstrative 'this man' fails to refer when it latches on to a cardboard cut-out because the conception internal to the demonstrative is that of a man, but there is no man. We can see this clearly in Anscombe's case of 'poor Jones':

> "Someone comes with a box and says "This is all that is left of poor Jones." The answer to "this what?" is "this parcel of ashes"; but unknown to the speaker the box is empty." (Anscombe, 1981, p. 28)

The conception internal to the demonstrative in 'this is all that is left of poor Jones' is 'this parcel of ashes'; but there is no parcel of ashes, so the demonstrative reference to the parcel of ashes fails. The demonstrative latches on to the empty box, but the conception internal to the demonstrative is not 'this empty box', it is 'this parcel of ashes', so the internal conception and that which the demonstrative latches on to do not line up:

> "The referent and what "this" latches on to may coincide, as when I say "this buzzing in my ears is dreadful", or, after listening to a speech, "That was splendid!" But they do not have to coincide, and the referent is the object of which the predicate is predicated where "this" or "that" is a subject." (Anscombe, 1981, p. 28)

[47] See, for example Kaplan: "An incomplete demonstrative is not vacuous like an improper definite description. A demonstrative can be vacuous in various cases. For example, when its associated demonstration has no demonstratum (a hallucination)—or the wrong kind of demonstratum (pointing to a flower and saying 'he' in the belief that one is pointing to a man disguised as a flower)—or too many demonstrata (pointing to two intertwined vines and saying 'that vine'). But it is clear that one can distinguish a demonstrative with a vacuous demonstration: no referents from a demonstrative with no associated demonstration: incomplete. (Kaplan, 1989, pp. 490-491)

When the conception internal to the demonstrative and that which the demonstrative latch on to do not line up, we have a failure of demonstrative reference. Note that this is not an error due to misidentification. Misidentification occurs because of the inferential grounding of assertions with names as subjects. There is no such inferential grounding for an assertion with a demonstrative as subject. Indeed, the suggestion above was that the assertion with name as subject was inferentially grounded in an assertion with a demonstrative as subject, which assertion is not itself inferentially grounded at all.

The demonstrative in an assertion with a demonstrative as subject can fail to refer, but the same could not be said of "I" in an assertion with "I" as subject: if "I" is a referring term, then it cannot be that, in such an assertion, "I" fails to refer. With this thought in hand, we are now in a position to see that assertions of 'I believe that *p*' do not merely fail to exercise the epistemic capacity for observational identification in getting hold of their subject, but do not exercise any epistemic capacity for identification at all in getting hold of their subject. Assertions of 'I believe that *p*' do not furnish the subject with identifying knowledge of the one she is. It is a short step from this to the claim that assertions of 'I believe that *p*' exercise no more epistemic capacities than those exercised in the assertion '*p*'. Crudely, uses of 'I' exercise no epistemic capacities at all, and as such, the claim that using 'I', unlike using 'A', is no cognitive achievement comes back into focus.

3.4. Answering the Puzzle of Transparency

3.4.1. Two Notions of Content.

As it stands, I have given no story that satisfactorily dissolves the puzzle of Semantic Discontinuity. What I have said so far is that we can pull apart the idea of content and the idea of truth-conditions, via Anscombe's point. But Anscombe's point tells us that two assertions can have distinct content while having the same truth-conditions (while obeying the same truth-conditional reference rule); the assertions 'I believe *p*' and 'A believes *p*' obey the same truth-conditional reference rule, but Anscombe's point is that they are distinct in content. The puzzle of Semantic Discontinuity asks how it can be that two assertions made on the basis of the same epistemic capacity can have *distinct* truth-conditions. The suggestion above (in section 3.2.1.) was that content should be understood in terms of truth-conditions. But what Anscombe's point shows us is that there is an equivocation on the notion of 'content'. If we equate content with truth-conditions then 'A' and 'I' have the same content. But Anscombe's argument shows that this cannot be so. Instead, we need a finer grained notion of content that avoids the equivocation; we can make sense of a notion of content that is independent of the truth-conditions of the assertion; assertions with 'I' as subject obey the same truth-conditional reference rule as assertions with 'A' as subject, but they differ in content.

Marshalling Anscombe's point to make some distinctions in the notion of 'content' in play should allow the semantic discontinuity of the assertions '*p*' and 'I believe that *p*' to come into focus.

The content that is semantically continuous across the assertions '*p*' and 'I believe that *p*' is a *general* content, content that is had in virtue of the exercise of an epistemic capacity – content that would be available to anyone who exercised such a capacity. The semantically continuous content is general insofar as it does not matter who makes an assertion containing '*p*', or whether '*p*' is embedded in a context (such as a belief clause), the content of '*p*', the *general content* stays the same. Content is in the sense I am discussing *general* insofar as it is independent of context. The assertion '*p*' has the general content *p* in virtue of the exercise of epistemic capacities.

We should pause here to say a little about epistemic capacities. The idea of an epistemic capacity is something that is the same across subjects and are something that does not provide for differences in content that turn on differences between subjects. The deliverance of an epistemic capacity depends for its truth on nothing about the particular subject who exercises the capacity. The exercise of an epistemic capacity provides only what I have called general content. Epistemic capacities themselves are then in this sense general; they can be had (and thus exercised) by any subject, and their deliverances are likewise general in content.

Contrast this general notion of content with the content of the assertion 'I believe that *p*'. The content of this assertion goes beyond the general content of the embedded '*p*', and what this 'going beyond' amounts to is precisely what the solution to the puzzle of Semantic Discontinuity must explain. The Simple Account holds that the 'I believe that _' does not reflect the further exercise of epistemic capacities, and as such the capacity to attach it to assertions of 'p' is no cognitive achievement. The 'I believe that _' is not arrived at through exercise of any epistemic capacity not exercised in the assertion '*p*' and since the general content of the assertion is had in virtue of the exercise of an epistemic capacity, the 'I believe that _' adds no general content to the assertion. The general content of the assertion, the content that is had in virtue of the exercise of an epistemic capacity is the content *p*.

But this is just recapitulating what has already been said. We have not yet accounted for the 'I believe that _'. The suggestion is that the 'I believe that _' does bear content, but it is not content that is had in virtue of the exercise of an epistemic capacity. What the 'I believe that _' provides in the assertion 'I believe that *p*' is a contribution which figures in determining the truth-conditions in a way that conforms to the truth-conditional reference rule. The 'I believe that _' in effect shows that the assertion turns to how things are with the asserter rather than how things are with the world for

its truth. It is in this sense that 'I believe that _' adds no general content to the assertion (no content that is had in virtue of the exercise of an epistemic capacity), but still adds a relevant content.

This way, we can see how both the content 'p' and the content 'I believe …' figure in the assertions '*p*' and 'I believe *p*'. The assertion 'I believe that *p*' bears the general content '*p*', which is had in virtue of an exercise of epistemic capacities and the content of 'I believe …' which contributes to the truth-conditions of the assertion, by telling us that the assertion turns for its truth not to the general content which is had in virtue of the exercise of epistemic capacities, but to the condition of the subject of the assertion (i.e. via the truth-conditional reference rule 'the assertion 'I believe that *p*' is true iff the speaker believes that *p*'').

The distinction drawn between the general content of *p* in the assertion '*p*' and the assertion 'I believe that *p*' and the non-general content of 'I believe _' is enough to defuse the Semantic Discontinuity puzzle; we can hold that the assertion '*p*' and the assertion 'I believe that *p*' are made by exercising of the same epistemic capacities but their truth-conditions are distinct by realising that there is a notion of general content that is had in virtue of the exercise of epistemic capacities, and a notion of content that is not, but which nevertheless contributes to determining the truth-conditions of the assertion. This moves us beyond the equivocation of content and truth-conditions. There is a notion of content that suggests that differences in contents are differences in epistemic capacities, and a notion of content that suggests that content is exhausted by truth conditions. But these need not be equivocated. What Anscombe shows us is that we can separate these notions. We have the idea of a general content which is available in virtue of the exercise of an epistemic capacity and a non-general content that is not. It is this non-general content that gives the solution to the puzzle. The 'I believe that _' tells us precisely that the truth conditions are distinct; we can hold that the truth-conditions of the two assertions are distinct because their non-general content is distinct.

3.4.2. The Simple Account and False Beliefs

I drew attention to a problem at the end of section 3.3; if the subject *NN* can be wrong in her assertion 'I believe that *p*', in a manner that is distinct from that of her being wrong in her assertion 'p', then surely the former assertion cannot be made by exercise of the same epistemic capacities as are exercised in the latter, and surely the subject *NN can* be wrong in her assertion 'I believe that *p*'. We can hold on to the idea that the assertion 'I believe that *p*' and the assertion '*p*' are made on the same epistemic basis, even though there is room for *NN*'s assertion 'I believe that *p*' to be false by considering the following thought: there is a canonical epistemic basis for the assertion 'I believe that *p*', and that epistemic basis is the very same epistemic basis as for the assertion '*p*' – this is what the understanding of self-knowledge provided to us by the Transparency Remark tells us. It tells us

that the epistemic capacity exercised in the assertion 'I believe that *p*' is nothing more than the epistemic capacity to assert '*p*'. If the assertion 'I believe that p' is made on its *canonical epistemic basis* — which is to say: if it is an exercise of the capacity to answer the question 'is it the case that p?'— then the assertion 'I believe that *p*' is guaranteed to be true. It is guaranteed to be true *because*, if the assertion 'I believe that *p*' is made on its canonical basis, then the subject will be in a position to give a positive answer to the question 'is it the case that p?', and as such will believe that *p*. As such, the subject's assertion 'I believe that *p*' will be true. This is not to say that the canonical basis is the only basis for an assertion of 'I believe that *p*'. It's perfectly plausible to think that the basis for the subject's assertion 'I believe that *p*' might be some non-canonical basis, such as that she had too much coffee and some thought popped into her head, or the subject is deeply irrational, and her epistemic capacities systematically malfunction. But these are outlandish possibilities, the sort of possibilities that should not undermine the credentials of the Transparency Remark as a way delivering self-knowledge. After all, if we wish to talk in those sorts of terms, the subject's assertion 'I believe that p' is an assertion of a safe belief, because, as the basis of her belief is the canonical one, her belief could not easily be wrong (indeed could not be wrong at all) [48]. By talking of the canonical basis of the assertion (i.e., the canonical exercise of an epistemic capacity) we can hold on to a Transparency account.

3.4.3. The Simple Account and Privileged Access

This explanation in terms of the canonical basis of assertion also gives a tidy explanation of how the Simple Account can explain the Authority of self-knowledge (as discussed in section 1.1.1.1). A subject's beliefs about her own mental states are in better shape, epistemically, than those she has about the mental states of others, or about the world. If our subject *NN* asserts '*p*' on the canonical basis of that assertion, it is guaranteed that her assertion 'I believe that *p*' will be true, and it is guaranteed irrespective of *whether her assertion that p is true*. Her assertion '*p*' does not need to be true; it is not any more secure even on this canonical basis, but the Transparent nature of belief guarantees the truth of her higher order belief.

We can also see an explanation for a sort of Groundlessness of self-knowledge. Self-knowledge is groundless in two (related) ways. Firstly, it is no cognitive achievement. In asserting 'I believe that *p*', the subject exercises no further epistemic capacities than those exercised in asserting '*p*'. Secondly,

[48] There is room here to advocate for an even stronger position, a position in which the Moorean paradox is a true paradox of rationality – that the Transparency account combined with the idea of the canonical basis of the assertion 'I believe that *p*' and the assertion '*p*' being the same basis tells us something about what it is to be a rational agent. The thought would be that it is not possible to rationally assert 'I believe that *p*' *except* when the assertion is made on its canonical basis. As such, the belief in the assertion 'I believe that *p*' would not merely be safe, it would be the only rational belief one could hold.

the grounds the subject has for her assertion 'I believe that *p*' are nothing more than the grounds she has for her assertion '*p*'. There are no special grounds associated with self-knowledge, and her assertion of her belief can indeed be understood as groundless insofar as it does not have any special grounds. It is not, however, based on *nothing*. Recall that in section 1.1.1.2, Cassam charged groundless self-knowledge with rendering self-knowledge *insubstantial*, as if self-knowledge is groundless, it must be based on *nothing*, and that which is based on nothing cannot be substantial self-knowledge (and so should be rejected). Cassam's objection gets no grip on the Simple Account – the knowledge delivered by the Simple Account is not insubstantial, the assertion 'I believe that *p*' is groundless insofar as the grounds are nothing more than the grounds of the assertion '*p*'. There are no special grounds for self-knowledge.

3.4.4. Identification and Predication

If the assertion of 'I' does not exercise the epistemic capacity for identification, however, we seem to reveal a problem. If no object is identified, how can anything be predicated of an object? If a speaker asserts 'I believe that *p*' our standard understanding is that the speaker predicates of an object (the very object they are) the property of believing that *p*. This would be in line with what is suggested by the truth-conditional reference rule. But this, as suggested, doesn't make out the difference between 'A' and 'me'. So, if the assertion with 'I' as subject does not exercise the epistemic capacity for identification, what is saying 'I believe that *p*' *doing*? If 'I believe that *p*' is not a (self) ascription of a property to an object, what is it?

3.4.5. Returning to Self-Ascriptions – The Generality Constraint

If the assertion of 'I believe that *p*' is not an ascription of a property to an object, then how should it be understood? I suggested above that the epistemic capacities exercised in the assertions '*p*' and 'I believe that *p*' must be the same; making the assertion 'I believe that *p*' involves no more epistemic capacities than making the assertion '*p*'. But why not think that *S*'s assertion of 'I believe that *p*' is still a self-ascription even if it is made by exercise of the same epistemic capacities as the assertion '*p*'? We are now in a position to marshal Anscombe's discussion to suggest an answer. The thought that the 'I'-assertion is not an identification, exercises no epistemic capacity for identification, suggests that there can indeed be no ascription of a property to an object. The subject *S*'s assertion that (as she would put it) 'I believe that *p*' does not, from *S*'s position, ascribe a property (belief) to an object (the thing that *S* is). The ascription of a property to an object cannot be what the subject *S* is doing when she asserts 'I believe that *p*', for if she were doing that, there would be no distinction between the subject *S* who has mastered the first-person pronoun 'I' and the subject *R*, who has mastered the use of 'A', and asserts 'A believe[s] that *p*'. That 'I believe that *p*' is no ascription is no answer to the question of how it should be understood, of course. Indeed, if an assertion of 'I

believe that *p*' is no self-ascription (understood in terms of identification and predication), we are faced with a greater puzzle – such assertions do not obey The Generality Constraint[49]. If a speaker avers 'I believe that *p*', it seems that the concept of 'belief' in play in a 'I believe that p', as the Transparency account understands it, is not *general* in the way that Evans suggests – the concept 'belief' is not playing the same semantic role in the assertions 'I believe *p*' and 'John believes *p*'. In the case of the first-personal assertion, the 'I believe _' does not perform the role of ascribing a belief that *p* to anyone. In the third-person case, the 'John believes _' performs (or appears to perform) the semantic role of ascribing a belief to an object (John). Anscombe's point denies that the Generality Constraint is of general application.

3.4.6. Predicables and Predication

Anscombe's point, it seems, presents us with a problem of generality. As discussed in the previous section, if assertions with 'I' as subject do not implicate identifying knowledge of the subject of the assertion on the part of the speaker, then it is difficult to understand how the assertions can involve the ascription of a property to an object (i.e., a predication). The suggestion in that section was that in the case of the assertion 'I believe that *p*' no such ascription of a property to an object occurs, but at that stage, no positive story was given. To reiterate, the puzzle is this: If a subject *S* asserts 'John believes that *p*', she is ascribing the property of believing *p* to an object, John. But if *S* asserts 'I believe that *p*', she makes no ascription of any property to any object. The predicate 'believes' is not *general* – it has a different semantic role in the first and third person cases, and as such assertions of the form 'I believe that *p*' are not constrained by the Generality Constraint. The subject's competence with 'believes' in the first personal case says nothing about her understanding of 'believes' in the non-first-personal case. The predication of 'believes' is not *semantically continuous* across first and third person contexts. It is important to make clear here that the first/third person asymmetry is a traditional problem in self-knowledge which is usually tied to the phenomenon of authority[50]. That is, given that first-personal attributions of belief are taken to be authoritative in some sense, either epistemically or semantically, it must have different semantic properties in the first-personal case than in the third-personal case, and this difference needs to be explained. The semantic continuity puzzle suggested here is not motivated by concerns of authority but is rather motivated by concerns of the logical form of the assertions in question -- what the role of ascription and predication in first-personal cases is. There is, however, a way to approach a solution.

[49] "…if a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception." (Evans, 1982, p. 104)

[50] See, in particular, Davidson (1984)

Consider the following thought – in both the first-person case ('I believe *p*') and the third-person case ('A believes *p*') there is still a *predicate* 'believe', but in the first-person case, that predicate is not predicated of any object, whereas in the third-person case, 'believe[s]' is predicated of e.g., John. We understand the semantic properties of the predicate 'believe' when applied in the third person case 'John believes that *p*', and we also have an account of the understanding of the use of 'I' – that is, an understanding of the truth-conditional reference rule and the application of that rule. If 'believes' is understood as a predicate that is in the first-person case not predicated of an object, but in the third-person case is, we can preserve the semantic continuity of '; the predicate 'believes' plays the same *role* across both assertions. The predicate 'believes' in both cases has no general content, i.e., content that is had in virtue of the exercise of epistemic capacities. Nonetheless, the content of 'believes' contributes to the determination of truth-conditions in both cases.

An objector might think, however, that 'believes' is still not general. If in the assertion 'I believe that *p*', even if 'believes' is a predicable which is not predicated of an object, the difference in whether or not the predicable is predicated of an object is (on this objection) enough to have it play a different semantic role and as such be semantically discontinuous in the first and third-personal uses. The objector here is indeed on to something. What fixes the content of an assertion of 'p' is what the capacity delivers, whereas what fixes the content of an assertion of 'I believe that p' is merely whether the capacity is exercised by the one who makes the assertion. What the capacity delivers contributes nothing to the determination of the content of the predicate 'believes', in the first-person case, and the challenge is whether this is also the case in the third-person case. This is a real challenge, and one that the Simple Account does not fully answer.

A suggestion, which does not constitute a complete answer to the objection is the following: In the first-person case, i.e. in the case of the assertion 'I believe that *p*', the content is a matter of whether the subject who makes the assertion exercises the capacity to answer a question of the 'is it the case that *p*?' form, whereas the content of a third-personal assertion of the form 'NN believes that *p*' is a matter of whether the referent of the name in the subject position (i.e. NN) exercise the capacity to answer a question of the form 'is it the case that *p*?' This suggests there is some uniformity across the assertions as both concern the exercise of an epistemic capacity. There is, of course a difference. In the case of the assertion 'I believe that *p*', the epistemic capacity that the assertion concerns is the very capacity whose exercise puts the one who makes the assertion in a position to make the assertion in question, whereas in the case of the assertion 'NN believes that *p*' the epistemic capacity the assertion concerns is not the capacity whose exercise puts the one who makes the assertion in the position to make the assertion in question--the assertion rather concerns a different capacity possessed by the one the assertion concerns, the one who is the referent of the name in

the subject position of the assertion (i.e. NN). It is at least not obvious that this difference in the epistemic capacities concerned should mandate a difference in the content of the predicate 'believes' across first and third person cases. Although this is an area where this thesis does not fully answer the question (the argument of this thesis is focussed on the case of first-personal assertions, even if I am sensitive to the issue of the continuity between first and third-person assertions), the suggested solution at least softens the impact of the objection while noting that this is an area where further work is needed. Further, the suggested solution suggests a (second) re-emergence of the traditional notion of first-person authority (where one can be authoritative regarding exercises of the epistemic capacity for belief in one's own case but not in the case of others), for two reasons: First, in the first-person case, the assertion does not implicate identifying knowledge of the subject and as such is immune to certain sorts of error, but in the third-person case, the same immunity does not apply. Second, in the first-person case, the epistemic capacity the assertion concerns is the capacity that puts the one who exercised the capacity in question in a position to make the assertion, but in the third-person case this is not so. As such the unity of subject and capacity in the first-person case gives some reason to think that the first-personal assertion is authoritative.

Thus, I suggest that we have here (enough of) a guarantee of semantic continuity across first and third person uses to preserve some uniformity in the predicate 'believes', so the puzzle is, if not dissolved, at least ameliorated, while recognizing that there is further work to be done that goes beyond the scope of this thesis.

3.5. Objections and Replies

3.5.1. The Objection from Anti-Luminosity

Any account of substantive self-knowledge must at some point engage with Williamson's Anti-Luminosity argument[51]. Williamson's argument is aimed to show that we have (as he puts it) no 'cognitive home', that is, there is no "realm of phenomena in which nothing is hidden from us." (Williamson, 2000, p. 93) Williamson identifies the idea of such a cognitive home with the idea of a realm where all conditions within are *luminous*. Williamson defines a condition C as luminous when it meets the following conditional:

> "(L) For every case *a*, if in *a*, C, then in *a* one is in a position to know that C obtains." (Williamson, 2000, p. 95)[52]

[51] Most prominently featured in *Knowledge and its Limits*, chapter 4, and extended somewhat in chapter 5. (Williamson, 2000)

[52] Cases for Williamson are individuated by subjects, times and worlds (Williamson, 2000, p. 94). The individuation of cases plays an important role in Williamson's argument, but we need nothing more than this

The principle is relevant for our purposes Evans' remark prima facie appears to be an affirmation that *belief* is luminous. By answering the world-directed question, I am in a position to answer the self-directed question. By asserting '*p*', I put myself in a position to assert 'I believe that *p*'. Thus, it might seem that if the anti-luminosity argument gets any grip, then an account of self-knowledge which is a development of the Transparency Remark must be false. After all, the argument Williamson provides suggests *no* interesting states are luminous, and my beliefs are certainly an interesting state.

3.5.1.1. The Argument and Initial Responses

The first response the supporter of Transparency can muster is to suggest that Williamson denies not just the idea that beliefs are luminous, but the thought that Transparency is supposed to elicit, that the assertion '*p*; and the assertion 'I believe that *p*' are made by exercise of the same epistemic capacity[53]. If one accepts that the assertions are made by exercise of the same epistemic capacity, luminosity (or something like it) should not be surprising, and the anti-luminosity argument gets no traction.

To see this, let's begin by reproducing the anti-luminosity argument Williamson provides for 'feeling cold', and recall that 'feeling cold' is intended to be a paradigm of a luminous condition:

> "(1_i) If in a_i one knows that one feels cold, then in a_{i+1} one feels cold.
>
> (2_i) If in a_i one feels cold then in a_i one knows that one feels cold.
>
> Now suppose:
>
> (3_i) In a_i one feels cold
>
> By modus ponens, (2_i) and (3_i) yield this:
>
> (4_i) In a_i one knows that one feels cold
>
> By modus ponens, (1_i) and (4_i) yield this:
>
> (3_{i+1}) In a_{i+1} one feels cold
>
> The following is certainly true, for a_0 is at dawn, when one feels freezing cold:
>
> (3_0) In a_0 one feels cold.
>
> By repeating the argument from (3_i) to (3_{i+1}) *n* times tor ascending values of *i* from 0 to *n-1* we reach this from (3_0):
>
> (3_n) In a_n one feels cold

coarse individuation. Further, nothing in the response I offer to the anti-luminosity argument turns on individuation of cases.

[53] Not all accounts that trade on Evans' remark unpack Transparency in this way. Those accounts are still hostage to Williamson's argument, but this proposed response will not work for them.

> But (3_n) is certainly false, for a_n is at noon, when one feels hot." (Williamson, 2000, pp. 97-98)

The motivation for the first premiss is that one knows in case a_i because the belief is formed on a reliable basis, and if the belief is formed on a reliable basis, then in case a_{i+1}, by stipulation a very short time after case a_i one must still be cold, or the belief in a_i would not have been formed on a reliable basis and as such one would not know one feels cold. Note that there is no discussion of the basis upon which one knows that one feels cold, rather the assumption going in to the anti-luminosity argument is that it should not matter what the basis one knows one is in a condition is (as long as it is a reliable basis).

Premiss (2_i) is the luminosity claim, and Williamson's argument suggests that (2_i) is incompatible with (1_i) and (3_i), and (2_i) has not been argued for, so should be dropped. The real action for us is in the combination of premiss (1_i) and (2_i). The suggestion here is that the reliably based belief in (1_i) leads to the claim about how one feels, which from (2_i) leads to a claim about one's knowledge of what one feels, and as (1_i) is a conditional, from a starting point where one feels cold, we have a sequence of steps that apparently lead to an inconsistent conclusion as there is a step somewhere in the sequence where one knows that one feels cold at a_{i-1} but does not feel cold at a_i, leading to the inconsistency at (3_n). This all turns on the assumption that if one knows one feels cold at a_i it must be the case that one feels cold at a_{i+1}, and this, perhaps, can be resisted by the supporter of Transparency. In order to make the sorities series the anti-luminosity argument provides work it seems that it must be the case that one's knowing that one feels cold is a distinct condition from one's feeling cold. If this were not so, it does not seem that there could be the 'epistemic gap' that the anti-luminosity argument needs to exploit to make headway — the gap between the condition and the knowledge of that condition. If there was only one condition here, there would be no distinction between the requirements that hold of one's knowing one feels cold, and the requirements that hold of one's feeling cold. But to suppose that there is no such distinction is to suppose that the requirement of reliability, which according to the Anti-Luminosity argument holds of knowing that one feels cold, as such holds of one's feeling cold. And then it seems that (1_i) could be re-written as:

($1`_i$) If in a_i one feels cold, then in a_{i+1} one feels cold.

This is obviously false, and yet it follows from the conditional Williamson set up, relying on the idea of a 'reliable basis' for knowledge, and the claim that there is only one condition here. The obvious response for the friend of Transparency is to suggest that Williamson's conditional is flawed for this reason. This can be seen clearly in the case of belief. In this case, the Transparency Account holds that one exercise of one epistemic capacity, puts the subject in a position to answer both a question

about the world ('Is it the case that p?') and a question about belief ('Do you believe that p?). The condition that consists in the exercise of this capacity just is: one's believing that p. And this condition just is: one's being in a position to answer the question of belief--that is, it just is: one's knowing that one believes that p (that is to say, it just is one's having self-knowledge of belief). There is one condition here, not two. The conditional 'moves' from the second order state of knowing that one feels cold to the distinct first order state of feeling cold. But this is not how the friend of Transparency would understand matters here. If the general shape of the Transparency Account were applied to the present case, it would hold that the condition of feeling cold is not a distinct condition from knowing that one feels cold. Of course, the supporter of anti-luminosity will simply deny that there is only one condition here, and as such the Transparency theorist's objection itself gets no traction. But flat denial is of course no argument, and we have good reason to think the Simple Account is a good account of self-knowledge of belief. (Whether we should apply the general shape of the account to the particular case that concerns Williamson is not a question we need take a stand on here. But we will return to this matter when considering the Objection from Scope.)

Further, recall that the Anti-Luminosity argument is not an end in itself, but rather is in the service of the idea that we have no 'cognitive home': there is no realm where nothing is hidden from us. And there is a sense in which the Simple Account is an account of self-knowledge that suggests that there is no cognitive home. The Simple Account suggests what is known in self-knowledge is not a 'cognitive home', in the sense of a special realm of facts, knowable by the exercise of an epistemic capacity that is such as to afford knowledge of these and only these facts. The Transparency Remark, properly understood, denies this special realm so understood – self-knowledge, whatever else it is, is the result of an exercise of the very same capacities exercised in knowledge of the world. The very idea of a 'cognitive home' is predicated on the notion that there is a cognitive domain that can only be known by exercise of capacities distinct from the knowledge in other domains. Indeed, if the 'luminous' knowledge the Anti-Luminosity argument is targeted against is knowledge had by cognitive achievement, the Simple Account agrees with the Anti-Luminosity argument.

The luminosity conditional (L) above makes no mention of cognitive achievement or epistemic capacities, but it is clear from the Anti-Luminosity argument that the higher order condition targeted in the argument is had by an exercise of an epistemic capacity not exercised in the obtaining of the first-order condition. This is what premiss (1_i) claims. If so, the Simple Account can agree that there is no cognitive home, and yet still hold on to the idea that (L) says something true (since as noted (L) is neutral with regards to the exercise of epistemic capacities).

3.5.2. The Objection from Over-Intellectualization

Recall that in section 1.4.3. I suggested that any account of Evans' remark must engage with the objection that Transparency accounts over-intellectualize self-knowledge. The over-intellectualization objection suggests that Transparency accounts demand too much from the subject – that is, that they demand too much cognitively. Standardly, the Rationalist version of this objection suggests that 'making up one's mind' misrepresents the phenomenon of self-knowledge and undermines the immediacy and groundlessness of self-knowledge; making up one's mind is not immediate in the sense self-knowledge is taken to be. While this objection might have some force against Moran's 'making up your mind' account, I suggest it gains no purchase at all against the Simple Account of Transparency developed in this chapter. The meat of the objection is that a Transparency account asks *too much* of the subject – it asks them to do more, cognitively, than the objector thinks an account of self-knowledge should. The Simple Account is a Rationalist account of Evans' remark which says that cognitively, the subject does *less* than the objector demands. One of the central insights of the account is that there is no special epistemic procedure or capacity for self-knowledge. The subject does nothing more to gain knowledge of their own belief than they would do to gain knowledge of the world. The objector might respond that the Simple Account still over-intellectualises self-knowledge, because it still involves the subject's entitlement to reflect on their assertion '*p*' to move to the assertion 'I believe that *p*'. This I think misunderstands the nature of the entitlement to reflect. Charging the Simple Account with this objection is a result of tacitly assuming that moving from the assertion '*p*' to the assertion 'I believe that *p*' involves a further exercise of epistemic capacities. And this assumption is false. In asserting '*p*' the subject is already in a position to assert 'I believe that *p*'. They get there 'for free' epistemically. This is what the entitlement amounts to. In asserting '*p*', the subject is in a position to assert 'I believe that *p*'. There is no over-intellectualisation. In asserting 'I believe *p*', the subject does nothing more epistemically than they would in asserting '*p*'. As such the objection gets no purchase.

3.5.3. The Traditional Puzzle of Transparency

An objector might think that the discussion of semantic continuity and discontinuity above is no answer to the Puzzle of Transparency. To answer this objection, we should make clear the dialectical position the Simple Account has suggested. Recall that the loosely formulated traditional puzzle is the puzzle of how one capacity can answer two questions – the question of what is the case and the question of what I believe. The solution the Simple Account suggests is that one capacity can answer both questions because *they have the same content*, in a sense of 'content' on which content is determined by epistemic capacities. But this solution brings with it its own worry – the worry that the answers to the question 'is it the case that *p*?' and the question 'Do you believe that *p*?' surely

have different content, because they have different truth-conditions. This is the puzzle of Semantic Discontinuity. The answer to that puzzle is to suggest that the notion of 'content' is doing double duty – it is compatible with the answers having the same content in the sense of 'content' I have labelled *general content*, content that is determined by epistemic capacities, that they have different contents in the sense of 'content' in which content is not determined by epistemic capacities, but is rather determined by (e.g.) the fact that the answer is given by one person and not another.

The objection an interlocutor might offer is that this attempt to dissolve the puzzle of Semantic Discontinuity simply gives rise to a version of the original Puzzle of Transparency – how can the exercise of one epistemic capacity answer two sorts of questions, one with content in the first sense, and one with content in the second? Surely when understood this way, there is no objection here at all – why should it be mysterious that the exercise of one epistemic capacity answers two questions with two sorts of content when the second sort of content is not determined by epistemic capacities? Surely it would instead be mysterious if a second capacity were required to answer a question with the second sort of content? The interlocutor might respond that what is ultimately mysterious is that there are two notions of content at all. This, I suggest, would be dialectically unappealing. As things stand, the interlocutor is objecting to the dissolution of the Puzzle of Semantic Discontinuity but is doing so by appealing to a reconstructed version of the Puzzle of Transparency. But this does not attack the motivation for thinking there is content that is determined by epistemic capacities and content that is not, i.e., Anscombe's point.

Dialectically, this should defuse the objection – the challenge is to make plausible the two sorts of content needed to dissolve the reformulated Puzzle of Transparency in section 3.2., the puzzle formulated in terms of distinguishing the truth-conditions of the assertions '*p*' and 'I believe that *p*' if they are made on the same epistemic basis. The appeal to general and non-general content is motivated by concerns independent of the Transparency Remark, namely by Anscombe's concerns with the semantics and epistemology of the first-person (although the ultimate formulation in terms of epistemic capacities is, of course, indebted to the Transparency Remark). As such, to challenge the intelligibility of the appeal to two kinds of content, the objection is not the puzzle of Transparency but a challenge to Anscombe's point. This is a substantive answer to the Puzzle of Transparency insofar as it answers the question of truth-conditions, and also deals with the question of why we might think there are two sorts of content. What this answer does not do is give a complete specification of those two sorts of content – all that is needed to answer the Puzzle of Transparency is a distinction between content that is had in virtue of the exercise of epistemic capacities and content that is not.

3.5.4. The Objection from Scope

As presented, the Simple Account of Transparency, insofar as it presents an account of self-knowledge at all, presents only an account for belief (and an interlocutor might suggest, only *occurrent* belief). The objection here is that we have self-knowledge of a wide variety of states, from our desires to our sensations, not just our beliefs, and a suitable account of self-knowledge should be (as Byrne (2018) suggests) *economical* – it should explain all self-knowledge in one explanatory swoop. This objection is, I suggest, the most pressing objection to the Simple Account. Answering this objection is the central focus of the next chapter.

3.6. Concluding Remarks to Chapter 3

We can now see just how radical Evans' insight is. Evans' insight can be taken to be that one comes to self-knowledge not through some special mechanism (an inner sense, say) or method (introspection, for example), but in virtue of *exercising the epistemic capacity to know at all*. Self-knowledge is had simply in virtue of the exercise of the epistemic capacities used in knowledge of the world. Further, the Transparency Remark is also in a sense radically deflationary about self-knowledge. If the Transparency Remark is indeed true, then self-knowledge has no special topic, in that it does not posit a "cognitive home" in the sense pinpointed above. There is no *cognitive achievement* associated with transparent self-knowledge. In this way, we also make sense of Boyle's remark that "The reflective approach thus does not seek to explain how we acquire doxastic self-knowledge." (Boyle, 2011, p. 6) Doxastic self-knowledge, ordinarily understood. would be a cognitive achievement on the part of the knower over and above their knowledge of how things are. The Transparency Remark offers no such thing. Instead, we are offered an account of self-knowledge of belief which asks for nothing more of the believing subject than that she be a believer. But if this is the case for belief, what of other domains of self-knowledge, such as the subject's self-knowledge of her intention, or her self-knowledge of her sensations? To ask this is to point to the one central objection which still remains, the Objection from Scope. In the next chapter, we will engage in depth with the Objection from Scope, and, in responding to it, the prospects for a general account of self-knowledge based on the Transparency Remark will become clear.

4. The Objection from Scope: Two Kinds of Self-Knowledge?

A central objection to Transparency accounts of self-knowledge is the Objection from Scope – i.e., the objection that accounts of self-knowledge which are based on or try to develop Evans' remark are too limited in scope to be satisfactory accounts of self-knowledge[54]. Matthew Boyle's (2009) *Two Kinds of Self-Knowledge* attempts to engage with Richard Moran's (2001) Rationalist account of Transparency in such a way that it shows the Objection from Scope to be misguided. In *Two Kinds of Self-Knowledge*, Boyle aims to fulfil three related aims:

1) To draw out that which is good/correct in Moran's 'making up your mind' account of Transparency[55], while acknowledging the criticisms levelled against Moran's view.
2) To understand in what sense the account of self-knowledge Moran develops is *fundamental*
3) To suggest that, in the face of the criticisms of Moran, we are forced to conclude that there is no *unified* account of self-knowledge to be had.

In fulfilling these three aims, Boyle also sets himself a fourth:

4) To give an account (or minimally, a sketch of an account) of how the capacity for the sort of self-knowledge Moran takes to be fundamental is a condition on the power to have thoughts with complex contents.

This chapter will aim to understand Boyle's engagement with Moran, and in doing so will aim to show why, once we have in focus a proper understanding of the Transparency Remark, we can see why the self-knowledge Evans' remark picks out is indeed *fundamental* to our cognitive lives, and why Boyle's third aim, and consequent acceptance that there are 'two kinds of self-knowledge,' is misguided.

To do this, the chapter will proceed in three stages. In the first section, I will give an account of how Boyle understands Moran's position[56]. In the second section, I will discuss how Boyle's account falls short of providing an account of self-knowledge by his own lights, drawing on discussion from Anscombe's (1981) *The First Person*. I will also discuss how we might rehabilitate the view Boyle puts forward to be in line with Anscombe's point in *The First Person*, and how this might relate to Boyle's fourth aim. In the third section I will engage with the question of fundamentality and unification

[54] This objection appears to be primarily aimed at the rationalist accounts of Transparency, but it is certainly a concern that Alex Byrne is sensitive to, given his focus on giving a unified, economical explanation of self-knowledge in terms of a transparent inference. We will engage further with Byrne in chapter five
[55] The central representation of Moran's account is found in *Authority and Estrangement* (Moran, 2001).
[56] I will not, however, quibble with the accuracy of Boyle's representation of Moran, except where differences lead to a clarification of Boyle's position. The goal of this exercise is to understand Boyle's position on its own merits, not merely as a representation of Moran.

(what I take to be Boyle's second and third aim), and how Boyle's explication of Moran's position brings out an important constraint on a satisfactory account of self-knowledge, which in the end Boyle's own development of Moran's view fails to respect. I will also suggest here that Boyle's suggestion that a unified account of self-knowledge is not within our grasp is mistaken by giving a general account of self-knowledge which is motivated by the Transparency Remark and would serve as the basis of a truly unified explanation of self-knowledge.

4.1. An Outline of Moran's Position

Boyle summarises Moran's position like this:

> "[Moran] argues that our ability to know our own current beliefs, desires, and other attitudes can on at least some occasions be understood as reflecting an ability to "make up our minds:" an ability to know our minds by actively shaping their contents." (Boyle, 2009, p. 134)

To understand what it is to 'know our minds by actively shaping their contents', we should have an idea of what Moran's explanatory goals in offering such an account are. Moran aims to offer a solution to 'The General Problem of Self-Knowledge', the problem of

> "…explain[ing] how we can be in a position to speak about our own minds in […] an immediate and authoritative manner, while still counting as speaking about the very same states that can be known to others only on the basis of observation or inference." (Boyle, 2009, p. 136)

Here *immediacy* is understood in the same manner as *Groundlessness* in section 1.1.1.2 (i.e., the claim that self-knowledge need not rely on or be based on evidence), and authoritativeness should be understood in the same way as A*uthority* in section 1.1.1.1 (i.e., that ascriptions of self-knowledge have better epistemic credentials than ascriptions of knowledge to others). We might think of this as a problem of the *semantic continuity*[57] of our attributions of belief[58] to ourselves and our attribution of belief to others: How is it when I say 'I believe that *p*', in self-knowledge (i.e., groundlessly and with authority), I mean the same thing by 'believe' as when I say 'John believes that *p*', based on observation or other evidence?

Moran suggests that this general problem really comprises two problems, one related to *attitudes*, and one related to *sensations*. Moran's focus is on the problem as it relates to *attitudes*. So, Moran's focus is on the question of how each one of us can make knowledgeable reports of our own attitudes which are authoritative and groundless yet semantically continuous with our attitude attributions to others.

[57] I take this idea to come from Davidson (1984), although not in those terms.
[58] And therefore, attributions of knowledge

Boyle suggests that Moran begins to answer this question by making the following observation concerning our attitudes: It seems that we can, in very many cases, come to know whether we hold certain attitudes by deliberating about the topics they concern;

> "If I want to know whether I believe that *p*, it seems that I can normally answer this question by considering whether there is reason to believe that *p* - whether there are persuasive grounds for thinking that *p* is true." (Boyle, 2009, p. 136)

This, suggests Moran, is true for *intending, hoping, fearing, desiring* and so on. Although there are cases where this is not true (recalcitrant attitudes, for example), the general thought is that it is striking that in very many cases, we can know our own attitudes by deliberating about the topic of the attitude in question. This striking observation itself raises a further question, a question which takes centre stage in discussions of Evans' Transparency Remark: how can it be that one answers a question regarding one's own beliefs by answering a question regarding ostensibly independent states of affairs in the world?

Boyle answers:

> "Self-knowledge of attitudes involves a sort of *agency*: I can know whether I believe that *p* by deliberating about whether *p* because my deliberation about *p* can constitute my making up my mind to believe that *p*." (Boyle, 2009, p. 137)

This is supposed to work like this: Upon being asked 'do you believe that it is raining?', I answer the question by reflecting on the reasons in favour of an answer concerning the weather. This reflective process would provide an answer to the question regarding my belief about the rain only if I can assume that the reflection on reasons to reach a conclusion determines what my belief is[59].

Our ability to speak authoritatively about immediate beliefs is, suggests Boyle, made intelligible

> "…if we suppose that to conclude that p on the basis of deliberation normally just amounts to coming to believe that *p*, and that a subject who possesses the concept of belief will understand that this is so." (Boyle, 2009, p. 137)

In a slogan, one can answer the question of whether one believes that *p* by reflecting on reasons or grounds for taking *p* to be true.

However, properly to understand what Boyle takes to be happening in Moran's proposal, we need to understand a minimal condition on self-knowledge and how this minimal condition delimits the discussion.

[59] Note that if one thinks that one's avowals of one's own beliefs do not exhibit authority or immediacy, then this view need not be compelling – one could infer from behavioural evidence and so forth that one has certain beliefs, but this would not be immediate or have first person authority.

4.1.1. A Minimal Condition on Self-Knowledge

Boyle suggests that the following is a minimal condition on a subject *S* expressing their knowledge:

> If a subject *S's* utterances are to count as expressing knowledge, *S* must understand whatever sentences she uses to express this knowledge.

This minimal condition entails a further basic requirement on self-knowledge: to understand the sentences she uses, "…a self-knower must *represent* her own condition as being of a certain kind." (Boyle, 2009, p. 143). It is by providing an understanding of both why the minimal condition is a justified condition on self-knowledge, and why this minimal condition involves the epistemic subject *S* having a particular sort of self-representation of her condition in knowing that we will see how Moran's account has philosophical import.

4.1.2. The Trained Parrot

To see how the capacity for self-knowledge depends on self-representation, consider the contrast between, on one hand, a competently self-ascribing speaker who expresses self-knowledge and on the other, a trained parrot.

The parrot is trained to cry out 'I am in pain' only when it is, in fact, in pain. It seems plausible to say that this utterance is not made on the basis of inference or observation. But it also seems to be the case that that the parrot *does not understand what it is saying*:

> "…it utters a form of words with a certain conventional content, but it does not grasp this content. And this implies that the parrot's vocalizations cannot express knowledge of its own pain in the way that similar sentences might in the mouth of a competent speaker." (Boyle, 2009, p. 143)

The parrot's utterance is surely a learned addition to whatever behaviours parrots have for expressing pain, an addition to the natural expressions of pain that parrots have. Surely this extension of a natural expression of pain does not manifest the parrot's knowledge that it itself is in pain, but merely manifests the pain. Ascriptions of knowledge are reserved for cases in which the creature acts in a way which manifests not just the natural expression of pain, but a grasp that it, the creature itself, is the very one who is in pain – a grasp that a certain subject is in a certain condition.

> "For whatever else it requires, knowing that *p* presumably requires representing that *p*, and only where a creature's activity expresses the attribution of a certain property to a certain subject is there a ground for saying that it has any representation of its condition at all." (Boyle, 2009, p. 143)

A subject who sincerely and competently avows 'I am in pain' must understand what 'pain' is and take herself to be in such a state – she *takes it to be the case* that she is in pain. The alternative is to suggest that we do not have authoritative knowledge of our own mental states:

> "… it would amount to reclassifying the statements that apparently express self-knowledge as mere automatic responses, which perhaps entitle an observer, or the subject himself, to judge that he is in a certain mental state, but which do not themselves express such a judgment." (Boyle, 2009, p. 144)

There is no semantic continuity between the utterances of a competent speaker who exhibits understanding of the language they speak and the utterances of the well-trained parrot. Insofar as the parrot's utterances of 'I am in pain' are contentful at all, they do not bear (or express) the same content as an understanding utterance of 'I am in pain'. The parrot's utterance is a 'natural expression' in the expressivist sense, which despite having the surface form of a subject-predicate expression does not predicate anything of any subject[60]. The parrot's utterance does not predicate anything of any subject because it does not bear the content of a speaker's saying that they themselves are in pain (in the same way that a moan or groan would not say bear that content). The parrot's expression is not a predication of a property to a subject; it is a manifestation of a disposition to behave in a certain way. Despite their surface similarity, the understanding utterance expresses content that the parrot's utterance does not.

To take this difference seriously, we need to (following Boyle (2009)) differentiate two different modes of expression of mental states:

> **Expression$_M$**: The Manifestation Sense, the sense exemplified by the parrot, where the expression of a mental state merely manifests that state.
>
> **Expression$_R$**: The Representation Sense, the sense exemplified by the competent speaker, where the expression of the mental state represents that mental state as one had by the speaker.

The subject exhibits self-knowledge only in cases where the speaker R-Expresses her mental state. If the speaker M-expresses her mental state, she does not represent herself as being in that state and hence the utterance cannot amount to knowledge that she is in that state. In the case of M-expression, the subject's behaviour is sufficiently accounted for by appeal only to the state in question, not to the subject's knowledge of her own state[61]. So, Boyle's claim is that for an utterance

[60] Note that this is unlike the Simple Account, which suggested that assertions with 'I' as subject do not predicate anything of a subject. The suggestion there was that the assertion contained a predicable which is not predicated. In the case of the parrot, it seems apt to say the sentence does not even contain a predicable – the parrot does not merely fail to represent itself as being in a certain condition, it does not represent anyone being in any condition at all. It merely manifests a condition.

[61] Note that Boyle does not restrict this to linguistic behaviour: "What is crucial is not that the creature should express its self-knowledge in an articulate language but that whatever sort of activity is supposed to manifest this knowledge should have a certain kind of explanation: one that adverts, not merely to the creature's being in the mental state supposedly known, but to the creature's representing its own state as of a certain kind." (Boyle 2009, p. 144)

to express self-knowledge, the uttering subject must represent *herself* to be a certain way. But this raises the question: what does such self-representation amount to?

Firstly, the kind of representations Boyle has in mind are *personal level* representations, not the sort discussed in a psychological theory of e.g., computational tasks in the brain. They are representations that are candidates to be available to the subject[62]. That is, these are self-representations which are "…predicable of the subject herself" by the subject herself (Boyle, 2009, p. 148), and as such each is a candidate to be a representation of the subject's mental state *as her own mental state*. These sorts of representation are (in a subject with relevant linguistic competence) paradigmatically employed in an utterance involving (a) subject-predicate form and (b) the form of the first person, e.g. 'I am F'. Further to this, a subject's utterance of 'I am F' R-expresses her condition only if her producing of the token utterance reflects her understanding of the meaningful elements of which the utterance is composed. An understanding utterance of 'I am in pain' produces the sentence in such a way that her use of 'I' expresses a comprehending representation of herself as the subject of the utterance, and her use of 'am in pain' comprehendingly represents that she predicates a certain state of that very subject.

To understand an utterance is to understand it as a complex of elements (at least, for example, as comprising a self-representation and a representation of an attitude), and for each of those elements to be meaningful, both independently and taken as a whole. So, a meaningful utterance is a complex representation composed of meaningful elements. When a subject R-expresses an attitude, that subject makes an utterance which she, the uttering subject, understands. To make an understanding utterance of the present kind, the subject represents *herself* in a certain way, and this representation has a particular form; the form of the first person.

4.1.3. Understanding R-Expression

So, to understand what it is to R-express an attitude, we must answer two related questions:

1) What is involved in understanding something as a complex representation composed from meaningful elements?
2) What is involved in understanding the particular elements that figure in a *self*-representation?

[62] This availability may not be easy, of course – it might require significant cognitive effort to become aware of e.g., one's recalcitrant attitudes, but nevertheless, this use of 'representation' targets a different phenomenon than the psychological theorist's use.

4.1.3.1. Understanding Complex Representation

To answer the first question, at least part of what it requires to understand one's utterance as a complex representation composed of meaningful elements is to be able to reflect on how the content of a given sentence relates to the content of various other sentences. To suppose that a token utterance of 'I' reflects general competence with the first person supposes that the use of 'I' in the sentence has a certain sort of explanation, one which points to the general abilities to competently employ certain linguistic tokens in sentences in general. And the same goes for predications such as '…is in pain'. Further, the subject must be able to recognize the connection between the truth of any one of these sentences and the truth of any other. This is a version of Evans' Generality Constraint[63], with a further capacity implicated, namely that of the ability to *reflect* on the relationships between contents of the complex representation and the contents of other complex representations. These relations are the relations of logic – e.g., exclusion, union, negation, and logical implication. In understanding the generality of sentences, one understands that they exist within a system of logical relations to other sentences. The sentences 'A is in pain' and 'A is not in pain' exclude one another, for example.

Boyle, however, suggests that the relations understood by the subject who is entitled to make complex utterances are more than the logical relations between sentences, and utterances of sentences – he suggests that the subject also understands support relations:

> "For to suppose that her knowledge of what it is for a subject to be in pain, or of what it is to ascribe a property to herself, is exercised in her understanding of these various sentences is to suppose that her understanding of them depends on her recognizing a common element in them, and recognizing this element as common must involve the capacity to recognize relationships (e.g., of implication, exclusion, and *inductive support*) between their contents." (Boyle, 2009, p. 150) [emphasis mine]

Boyle specifies 'inductive support', presumably understood as 'supporting a generally good inductive inference'. That the sentence (x) 'It has rained for the last thirty minutes' supports a good inductive inference to the sentence (y) 'It will rain for another ten minutes' is not a logical relation between (x) and (y). That (x) is the case does not logically entail (y)'s being the case. Rather (x) speaks for or supports (y). When asked why one believes (y), one points to (x); (x) is a reason or ground for (y). Boyle here moves from an abstract description of relations between sentences to a description of relations between beliefs. Sentences are the wrong sort of thing to be related by support relations. Sentences are only supported or unsupported when they express beliefs – what is supported or unsupported is not the sentence, but the belief that the sentence expresses. The move from

[63] "...if a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception." (Evans, 1982, p. 104)

relations between sentences to relations between belief contents parallels a move from talking of sentences to talking of 'claims' – that is, *asserted* sentences. The Generality Constraint constrains the possible contents of complex representations (i.e., complex sentences). Whether or not those sentences are asserted or claimed is independent of the constraint. The Generality Constraint says what it is for something to be a complex representation, not what it is for a subject *S* to make a claim with a complex representational content. This move from sentences to claims is where Boyle goes further than Evans, and why Boyle needs to invoke support relations as well as the relations of logic. As Boyle puts it:

> "A comprehending speaker, then, must be able to make claims in a way that reflects a grasp of the relation of the content of any given claim to the contents of a system of possible other claims." (Boyle, 2009, p. 150)

There is a move from a system of sentences to a system of claims. Claims are the sort of thing that are supported or unsupported by other claims, and claims are the sort of thing that are *made*. A subject makes a claim that such and such is so. By making a claim to this effect, the subject is moving into the Space of Reasons – the space of making claims and giving grounds for claims made. We might think of Boyle as offering a second constraint, one on the making of claims rather than on the possession of complex representations. Call this the Claiming Constraint:

> For a subject S to claim that *p* she must have the capacity to grasp the support and countering relations between the content of her claim and the content of any other claim

On the face of it, it may be difficult to see how Boyle warrants this move from a recognition that understanding complex relations require a certain generality of thought and a comprehension of logical links between representations to the idea that asserting certain sentences (themselves complex representations) requires a grasp of support relations. In essence, what Boyle is looking for is an *understanding-assent link*[64] as a central feature of self-knowledge.

We can perhaps see the motivation for Boyle's move by considering the generality claim in light of his suggestion regarding the understanding each subject displays of the relationship between the truth of her complex utterance and the truth of any other utterance. The recognition of relationships between the truth of a particular sentence-content implicates the recognition of how the truth or

[64] I take this term from Williamson (2007). Williamson's text argues against the possibility of such links. While this is a potentially important criticism of Boyle, I will only draw attention to the link between the two discussions here – a satisfactory treatment of Williamson's discussion is beyond the scope of this chapter. It suffices to say that if one is sceptical of the possibility of understanding-assent links then one should be sceptical of Boyle's proposal on that basis.

otherwise of such contents supports or undermines the truth of any other sentence-content[65]. Understanding the content of a complex sentence implies understanding its place in a larger scheme of sentence-contents, and the connections between them. If S understands the uttered claim 'I am in pain', she understands each part of the utterance as part of a larger whole, where the utterance 'I am in pain' supports other utterances, for example through implication. Take the further claim 'I need a paracetamol tablet', for example – the utterance 'I am in pain' supports the implication that one wants pain relief. As Boyle puts it:

> "… a subject could hardly be credited with the ability to grasp relations among various systematically related contents if her endorsement of any given content were not potentially open to modification by her consideration of its relation to other contents she endorses." (Boyle, 2009, p. 151)

Understanding this link between claims in turn requires that the subject be able to reflect on her reasons or grounds for holding a particular claim to be true. A subject must be able to ask a certain sort of question about the claims she makes – she must be able to ask *why* she takes the claim to be true. She must be able to give and ask for reasons.

Any subject who is capable of asking this sort of why-question will also be entitled to ascribe beliefs in the sort of way the deliberative account describes:

> "For a subject who can say that *p* just when she takes there to be sufficient grounds for supposing *p* to be true is a subject whose speech already expresses her beliefs: when she (nondeceptively) says "*p*," she will be affirming something she takes to be true, and since to take something to be true just is to believe it, she will also be entitled to say "I believe that *p*."" (Boyle, 2009, p. 151)

The ability to give reasons for one's beliefs already entitles one to self-ascribe beliefs in a way that reflects Evans' Transparency insight. The subject's utterance 'I believe that *p*' is an R-expression just in case *S* understands her utterance as a complex representation in which she represents herself as believing that *p*. Understanding her utterance as a complex representation in turn requires that *S* understand that her utterance is one amongst a system of claims, related to other claims through both the relations of logic and evidential support relations. Understanding these relations entails that *S* understands the grounds one would have for taking claims to be true. For a subject *S*, believing that *p* entails that *S* take *p* to be true, so for *S* not merely to believe that p but to represent herself as believing *p*, she must take *p* to be true (and represent herself as such). This in turn entitles *S* to reflect on why she takes *p* to be true – to ask a certain sort of why question: 'why do I take *p* to be true?', i.e., 'on what grounds or for what reason do I take *p* to be true?'. If *S* takes *p* to be true

[65] This does not imply that the truth of one sentence relates to the truth of every other, two sentences could happily be isolated from each other in terms of their truth.

just when there are sufficient grounds or reasons for supposing that *p* is true, she is entitled to R-express her attitude, i.e., she is entitled to say 'I believe that *p*'. That is, the subject who has the ability to R-express her attitudes is entitled to move from the assertion '*p*' to the assertion 'I believe that *p*' on the basis only of the ability to R-express her attitudes. In asserting '*p*', *S* takes *p* to be true, and does so on the basis of reasons or grounds that she understands herself to have. She represents herself as believing that *p* when she asserts '*p*'. *S* is entitled to accompany her assertions with an 'I believe'. And this entitlement to accompany her sincere assertions amounts to the subject making up her mind that *p*, taking up the deliberative stance the deliberative account suggests.

4.1.3.2. Understanding Self-Representation

To answer our second question, namely 'What is involved in understanding the particular elements that figure in a *self*-representation?', we need to understand what figures in a subject *S* having the right sort of understanding of a *self*-representation that figures in a complex representation. For all that has been said so far, we only have it that a subject who has the power to R-express her mental states is *entitled* to accompany her sincere assertions with 'I believe,' not that she must grasp this entitlement[66]. But Boyle thinks the connection runs deeper.

Boyle's position is that a claim is a self-ascription in the interesting sense only if it involves a 'form of the first-person'. A claim is a self-ascription in the sense we are interested in (i.e., a self-ascription which could amount to self-knowledge) only if it has a certain form, the form expressed in the understanding use of the first-person pronoun and a predication (or ascription) of some property. The claim is a self-ascription if it has the form 'I am F', where the ascriber understands what the complex representation 'I am F' expresses and understands the components of the complex representation (i.e., the 'I am F' is an R-expression of the speaker's attitude). Our question is what is involved in the *self-representation* component of the complex representation 'I am F', and we have suggested that it involves a form of the first person. Boyle suggests that

> "An expression "*A*" is a form of the first-person only if a subject who understands it understands that, in saying "*A* is F" he is predicating the property of being F of himself, i.e., the very person who is claiming that this predicate applies to this subject." (Boyle, 2009, p. 153)

If the subject did not understand that in expressing an attitude, she predicates that attitude to the very subject she is, then her predication of an attitude would not be a *self-conscious* predication,

[66] Even if we grant that this is the case, Boyle's point about the fundamentality of the deliberative account comes into focus: "…this already implies that an account of self-knowledge must leave room for the possibility of the sort of self-knowledge that Moran describes, since it implies that any subject who can make comprehending assertions is one who could acquire self-knowledge through deliberation simply by coming to understand the relevant entitlement." (Boyle 2009, p. 152)

even in cases where she succeeded in referring to the one who believes *p*. If it is as Boyle suggested, and to be able to R-express a given claim requires in principle considerations of the grounds of the claim, then:

> "It follows that a subject's use of "*A*" will express self-consciousness only if it bears the right sort of connection to this ability: he must understand that the person he calls "*A*" is the very person whose mind is, so to speak, his to make up." (Boyle, 2009, p. 153)

The speaker must understand the connection between the one she calls '*A*' (the one she uses the first-person pronoun to pick out) and the grounds by which *p* is taken to be true. She must understand that in saying '*A* believes that *p*', *A* takes *p* to be true, the grounds on which *A* takes *p* to be true are the grounds had by *A*, and that she herself is *A*, so she herself takes *p* to be true, and does so on grounds she herself has. Her use of A must be a "self-conscious use", in this sense.

Contrast this with a subject who uses an expression, let it be "*A",* in a non-self-conscious form. If the subject utters '*A* is about to be hit in the head', while she retains the normal aversion to head injuries, but takes no evasive action, we can plainly see that her use of '*A*' is not the self-conscious use – she does not understand that she herself is *A*. As Moran suggests in the text Boyle's account is drawn from "…ordinary self-knowledge … provides reasons for action" (Moran, 2001, p. 67). The subject who uses *A* in the non-self-conscious form makes an utterance, but this utterance is not one of self-knowledge. The knowledge she has does not provide a reason for her to act. If she understood that she was *A*, she would try to avoid the head injury. Her actions and beliefs are not properly connected with her representation of the one who she is – she does not have the belief that (as we would put it) 'I am *A*'[67], so that *A* is about to be hurt is not a reason for her to move, because (to her knowledge) *she herself* is not about to be hurt. She does not take it to be true that she is about to be hurt.

Compare the subject who suffers an unfortunate head injury to the subject who is prepared to say 'it is raining' but needs to look for behavioural evidence to say that '*A* believes it is raining'. This subject does not connect her taking it to be true that it is raining and her taking it to be true that A believes it is raining. She does not understand that she is the very one who believes it is raining. She does not have immediate, authoritative knowledge of her own attitudes. Her knowledge of her own attitudes is based on observational evidence: she does not know 'straight off' what her attitude is, without having to check[68]. Further, it seems that, in this sort of case, where the utterance is not a

[67] This specific locution of the puzzle here is quite infelicitous since '*A'* is supposed, on this view, to 'stand for' or 'substitute for' 'I'. Nevertheless, the general point that the speaker does not understand the first-person pronoun and as such cannot self-represent is captured well enough for the purposes of the example.
[68] Her knowledge of her own attitudes is had by an exercise of an epistemic capacity beyond those implicated in her having of the attitude…

self-conscious one, the speaker would not speak with authority regarding her own beliefs. When she utters '*A* believes that *p*', she does so on the basis of observation of her own behaviour, as she doesn't connect up the fact that that she herself believes that *p*, the fact that *A* believes that *p* and the fact that she is *A*. So, she knows *A*'s attitude on the basis of observation of *A* (it just so happens that in this case she is *A*). But she could be wrong about *A*'s attitude, she might be confused, or hallucinating. She could also be wrong about who *A* is – we can imagine a case where her observational capacity goes wrong somehow and she associates the pronoun *A* not with herself but another, and she bases her utterance on the behaviour of the other, not herself. If she cannot use the first-person pronoun as an expression of her self-consciousness, i.e., to genuinely refer in the self-conscious way, she is, in some sense, alienated from her own attitudes.

We can now clearly see the connection between the self-representation and the adoption of the deliberative stance, the making up of one's mind:

> "[A] subject's use of "*A*" expresses self-consciousness only if he displays an awareness that his reaching the conclusion that *p* is the reaching of this conclusion by the thing he calls "*A*."" (Boyle, 2009, p. 154)

The self-conscious use of '*A*' connects the decisions the subject *A* makes with the intentional actions of the thing she calls *'A*'. Reaching conclusions in this sense is itself an action, a thing one does, and a thing one does on the basis of reasons or grounds. Possession of a self-representation of the sort entailed in the ability to R-express attitudes demands the ability to self-consciously use the first-person pronoun. If the subject who utters a sentence 'I am F', which R-expresses her attitude, she does so in line with the claiming constraint above; and this implies that she understands the component representations of the complex representation which constitutes her claim, and the relations of support her claim exists within. When she R-expresses her attitude (perhaps by claiming, as she would put it, 'I believe *p*'), the subject must be aware that it is *she herself* who has the very attitude that is expressed, if her expression is so much as to *be* an R-expression. She must understand that she represents herself (understood first personally)[69] as possessing a certain attitude.

> "…a subject who *does* understand that his affirming "*p*" is *A*' s affirming "*p*" displays an awareness that when he decides that a proposition is true, this is a decision of the thing he calls "*A*" […] his use of "*A*" reflects an understanding that his determinations of what is true are the determinations of the thing he calls "*A*"" (Boyle, 2011, p. 155)

If a subject *S* determines (i.e., makes up her mind) that *p* is true (based on some ground, reason or evidence, or nothing at all), then if *S* understands the first-person pronoun, i.e., has and understands

[69] That is, 'herself' here could also be marked with the herself* introduced by Castaneda (1966). Note that Boyle and Moran do not use Castaneda's formulation.

that she has a self-representation expressed by the first-person pronoun, she understands that she can affirm (for example) 'I believe that *p*'.

The upshot of this discussion is that it is a condition on the understanding of the first person at all that the subject have the power to make up her mind:

> "…if a subject does not possess a representation which is linked, in the sort of way just described, to his power to make up his mind about what is the case, then he does not possess the power of self-representation, and hence cannot entertain self-ascriptive thoughts. In particular, he cannot think thoughts about his own mental states. Hence he cannot be a self-knower." (p. 155-6)

Boyle's proposal links the capacity for self-representation and self-ascription to the capacity for self-knowledge, and this understanding of self-representation and self-ascription amounts to the capacity to deliberate—to make up one's mind that *p*.

4.2. Boyle and Anscombe

4.2.1. A Challenge for Boyle

Boyle's link between deliberation and self-knowledge relies on the thought that any device with a certain form behaves in the same way as the first-person pronoun[70] and that the understanding of this device, the power to use the first-person at all, is grounded or based in the ability to 'make up one's mind. The argument Boyle provides for the device '*A*' being (as he puts it) a form of the first person is in a sense an inversion of the strategy used by Anscombe in *The First Person* (Anscombe, 1981), where her strategy is to show how '*A*' and 'I' have the same truth-conditional reference rule, but are still importantly distinct. It is by reflecting on Anscombe's argument that we can see that it is not possible to understand '*A*' as a 'form of the first person' by understanding it as a device that satisfies Boyle's constraints.

To see this, we will proceed in two stages. First, we will carefully examine the constraint Boyle uses to motivate that '*A*' is a form of the first-person and suggest that the constraint leads to a circular explanation of what it is to be a form of the first person. Building on that, we will examine what Boyle says about his '*A*' to show that understanding '*A*' is based on understanding an identity of the form 'I am *A*'. Boyle's constraint is the following:

> "An expression "*A*" is a form of the first-person only if a subject who understands it understands that, in saying "*A* is F" he is predicating the property of being F of himself, i.e., the very person who is claiming that this predicate applies to this subject." (Boyle, 2009, p. 153)

[70] i.e., that '*A*' and '*I*' in the discussion above have the same properties, they are both 'forms of the first-person'.

This constraint says what it would be for an expression to have the form of the first person. It suggests that for an expression '*A*' to have such a form, it must be that when a speaker makes an assertion with '*A*' as subject that speaker understands something about that assertion, namely that in predicating something of the speaker of that assertion, she is predicating something of *herself*. We must be careful how we read the '*he*' and '*himself*' in the quote above (and the '*herself*' in the explication above). The natural reading of Boyle's constraint suggests that the pronouns following the that-clause 'understands that' in the constraint (namely, 'he' and 'himself') are already (as Boyle would have it) 'forms of the first person'. That is, they are Castaneda' special 'star' pronouns[71]. If this were the case, the understanding that '*A*' is a form of the first person would be had on the basis of the subject *already* employing forms of the first person, namely, those constituted by the 'he' and the 'himself' in "…he is predicating the property of being F of himself…" (Boyle, 2009, p. 153). If this is so, Boyle's constraint already smuggles in understanding of what it is for a device to be a form of the first person into the constraint on what it is for a device to be a form of the first person.

We might suggest that the understanding of the first-person Boyle requires is not of the sort targeted by Anscombe's argument – that Boyle instead suggests a form of understanding of 'a form of the first-person' that does not implicate identifying knowledge of the subject of the assertion 'A is F'. This, I think, is to misread Boyle. In explicating what it is to have a comprehending self-representation, Boyle makes the following claim

> "The kind of representation that is a condition of authoritative self-knowledge must be a self-representation in a stronger sense: it must be a representation of the subject's mental state *which is predicable of the subject herself* rather than merely of some hypothesized processing mechanism operating within her. But even this is not enough, for, famously, I can represent what is in fact my own state but fail to represent it *as* my own state. Thus, in John Perry's well-known example, I might represent that *somebody* is pushing a shopping cart containing a torn bag of sugar but fail to represent that *I* am pushing that shopping cart, even though I am in fact the person in question. The kind of self-representation that is a condition of authoritative self-knowledge must be a self- representation in a yet stronger sense: it must be a personal-level representation of the subject's mental state *as her own* mental state." (Boyle, 2009, p. 148)

There is much to unpack in this long quotation, but the central point I wish to draw attention to is the idea that the notion of self-representation Boyle contends is a condition on authoritative self-knowledge has the following feature: in representing herself as being a certain way, the subject knows something about herself. If she did not, the challenge from Perry's forgetful shopper would get no grip. The idea of Perry's forgetful shopper is that I can represent *someone* as being a certain way, but not know that the *someone* I am representing as being a certain way is *me*. That is, I do not

[71] See (Castaneda, 1966)

have the right sort of knowledge of the subject of the claim that 'someone is a certain way', so I cannot move from 'someone is a certain way' to 'I am a certain way' (via the proposition 'I am the 'someone' in question'), but I can get such knowledge by identifying myself as the 'someone who is a certain way'. Self-representation for Boyle is a matter of identifying oneself as the bearer of a predicable state[72]. Identifying oneself as the *bearer* of a predicable state is identifying oneself as the object predicated of, i.e., it is to exhibit identifying knowledge of the subject of exactly the sort that is Anscombe's target.

The objection from Anscombe's point is that if the understanding of '*A*' is based on understanding the identity statement 'I am *A*', then '*A*' simply is not a form of the first person. We can see the reliance on an identity statement even more clearly in Boyle's further explanation of the constraint:

> "It follows that a subject's use of "*A*" will express self-consciousness only if it bears the right sort of connection to this ability: he must understand that the person he calls "*A*" is the very person whose mind is, so to speak, his to make up." (Boyle, 2009, p. 153)

Recall that in section 3.4. we discussed Anscombe's point in relation to the Simple Account of Transparency. The contention, drawing on Anscombe's discussion was that assertions with 'I' as subject make no identification of that which is in the subject position *at all*, on pain of misidentification through error, or of reference failure. But Boyle's suggestion here is that the use of '*A*' expresses self-consciousness (i.e., is a form of the first person) if it is appropriately connected to an ability – the ability to understand that *the person he calls 'A'* is the very person he is. Note that Boyle's description here amounts to an understanding of '*A*' as a special sort of name, the name the speaker calls the person 'whose mind is his to make up'. It seems to be saying that the use of '*A*' expresses self-consciousness only if the subject's understanding of '*A*' involves knowledge of the identity 'I am *A*' (or perhaps '*A* is the name I use for myself')—where 'I' in this identity-statement is itself a form of the first person, and so itself expresses self-consciousness. Once again this is no explanation of what a device that is a form of the first-person amounts to, since the explanation given, as before, is circular. That the understanding of '*A*' rests on understanding an identity-statement is a more significant problem for '*A*' being a form of the first person.[73]

4.2.2. Escaping the Puzzle – the Simple Account and Reasons

Nevertheless, we should aim to hold on to the thought that Boyle gets *something* right. The basic insight that there is something right in saying I come to know what I believe by making up my mind

[72] This is captured in the discussion thus far as 'being a certain way' which is naturally understood as exhibiting a certain physical property, but it is not limited to that – see (Boyle, 2009, p. 148, fn. 18)

[73] Anscombe's discussion, in showing that understanding '*I*' cannot amount to identifying knowledge, also shows that '*A*' cannot be a form of the first person if it is a special sort of name, which appears to be Boyle's suggestion, adding further argument that his position is untenable.

about the topic of my belief is worth holding on to, as is the thought that self-consciousness and self-knowledge are tightly bound up. We can, I think, hold on to these thoughts by recognizing that while the distinction between M-Expressing and R-Expressing do delineate categories of expression, where we go wrong is in assuming that when a speaker R-Expresses a belief, the idea of her self-representation involves a self-ascription (and inter alia identifying knowledge of that which is ascribed to). This is familiar territory. In Section 3.4.6, I suggested that the Simple Account of Transparency entails that assertions of 'I believe that *p*' are not self-ascriptions (understood as predications of a property to a subject) of belief, on pain of involving identifying knowledge of the subject of the assertion.

We can preserve Boyle's insight that R-expressing the belief that *p* (by asserting '*p*') *entitles* one to assert 'I believe that *p*'. The Simple Account gives a story as to why that is so – the canonical R-expression of one's belief that *p* is the assertion '*p*', and the assertion '*p*' is made by exercise of the same epistemic capacities as the assertion 'I believe that *p*'. The entitlement Boyle suggests we have in virtue of R-expressing our attitude is an entitlement to move from the assertion '*p*' to the assertion 'I believe that *p*' which comes from the exercise of epistemic capacities, and the 'transparent' structure Evans draws our attention to. But Boyle's discussion places central focus on *self-representation*, and in particular the place such self-representation has as part of the power to have thoughts with complex contents. It is this relation between self-representation and the having of thoughts with complex contents which underlies Boyle's argument for the fundamentality of the sort of self-knowledge he suggests. Given that Anscombe's point strictly undermines Boyle's argument, we must recover this fundamentality point.

We should reflect on the distinction between M-expressing an attitude and R-expressing an attitude. The difference between M-expressing and R-expressing an attitude is, as Boyle rightly contends, the attitude's place in a larger complex of attitudes. M-expressions of attitudes are in a sense cognitively isolated. If the trained parrot M-expresses the belief that *p*, they can, as it were, go no further. The belief that *p* cannot, for example, be employed in an inference to *q*, the M-expressing speaker has no entitlement to do so; if the parrot expresses the attitude 'Polly wants a cracker', the parrot is not entitled to infer 'Polly is hungry' on the basis of their M-expression[74]. R-expressions are not like this.

[74] This point I think undermines Boyle's argument against expressivism. He suggests that R-expression presents a challenge for theories of self-knowledge:

> "Any theory of self-knowledge will confront [a] challenge […]: it must leave room for an account of what it is not just to have mental states but to represent one's own mental states." (Boyle, 2009, p. 146)

Expressivists suggest that avowals of pain are learned additions to pain-expressing behaviours. But the suggestion is that such expressions M-Express pain, but do not necessarily R-Express pain. If self-knowledge requires R-Expression then the expressivist is in trouble because they do not have an account of how such

R-expressing an attitude carries with it the entitlement that the attitude expressed can figure in a complex of attitudes. Recall that when elucidating the conditions on R-expressing, Boyle suggests that complex representations have a certain structure, captured by the Generality Constraint (or something like it). Another way to put this point (or the genesis of this point) is that complex thoughts exhibit a special sort of structure, and that structure is one that admits to a certain generality, so when S thinks 'A is F', the ability to think 'A is F' is part of a complex of abilities with a general structure – to think that 'A is F', S must also be able to think 'B is F' and 'A is G', where A and B are subjects of thought and F and G are properties predicated of those subjects. As suggested above, this ability consists in the understanding of the logical relations between the sentences which express the token thoughts[75].

Boyle's own proposal goes further than this minimal suggestion on the structure of thought, however – he moves from logical relations to *evidential support* relations:

> "For to suppose that her knowledge of what it is for a subject to be in pain, or of what it is to ascribe a property to herself, is exercised in her understanding of these various sentences is to suppose that her understanding of them depends on her recognizing a common element in them, and recognizing this element as common must involve the capacity to recognizing relationships (e.g., of implication, exclusion, and *inductive support*) between their contents." (Boyle, 2009, p. 150) [emphasis mine]

It is not clear what warrants this extension of the Generality Constraint to include evidential support relations, and Boyle appears to conflate logical and evidential relations in his discussion. The conflation of the two is problematic for the following reason: the competency with logical relations implicated in the Generality Constraint give us an account of what it is to be a thinker of complex thoughts. For a subject *S* to think thoughts with complex representational content, she must be competent to say that her thoughts are not logically incoherent. The logical relations in question relate only the contents of the thoughts in question; two contents cannot be logically incompatible for *S* to think coherently. The Generality Constraint is a constraint on what it is to think coherently. The introduction of evidential support relations expands the relata in question. The evidential support relation is not merely a relation between thought contents, it relates those thought contents with how things are independent of the thinker. Evidential support relations tell *S* how well

---

expressions do more than M-Express, without weakening the analogy between natural expressions of pain, like crying, and linguistic expressions of pain, that expressionism trades on. But this *cannot* be the correct reading of e.g. Bar-On's expressivism (see e.g. (Bar-On, 2012)), where such natural expressions are not inferentially isolated. Indeed, it is not clear to me whether even the simple expressivist is guilty of what Boyle suggests, although it could be understood as a version of the Frege-Geach problem for ethical expressivism. This point is worth exploring further but is beyond the scope of this thesis. Rather the point is that Boyle's argument is not as clear a rebuttal to expressivism as he suggests it might be.

[75] In this he aligns his proposal for making sense of Transparency with Evans' own.

supported or otherwise the content of her thought is. They tell *S* whether it is *valid* to think the thought in question, not merely whether the thought is coherent. As Boyle rightly points out, this moves thought into the Space of Reasons. Of course, we might think that to talk about thoughts is already to be within the Space of Reasons, and thus Boyle's move is a justified one. We might argue that in thinking that *p*, we are in a position to give and ask for reasons,[76] that operating within the logical space of reasons is a precondition on having complex thoughts, but this is a substantive claim which is not argued for explicitly. There is an implicit argument for this claim, however. Boyle suggests that

> "…when she [the subject] (nondeceptively) says "*p*," she will be affirming something she takes to be true, *and since to take something to be true just is to believe it*, she will also be entitled to say "I believe that *p*."" (Boyle, 2009, p. 151) [emphasis and clarification mine]

So, believing that *p* is taking *p* to be true; believing the proposition 'it is raining' is taking the proposition 'it is raining' to be true. And 'taking to be true' *is* the sort of state that operates in the Space of Reasons – by taking *p* to be true, *S* is operating under some standard, some norm, and is then in the business of saying *why* she takes *p* to be true. In taking *p* to be true, *S* is making the judgement that *p* is so, so believing that *p* is *inter alia* judging that *p*, and judgement operates in the Space of Reasons.

4.2.3. An Echo of McDowell

It is here we see an echo of McDowell, i.e., the idea that

> "…the very idea of representational content, not just the idea of judgements that are adequately justified, requires an interplay between concepts and intuitions, bits of experiential intake. Otherwise what was meant to be a picture of the exercise of concepts can depict only a play of empty forms." (McDowell, 1996, p. 6)

And the further idea that the Given should not get a grip on our discussion of thought:

> "The idea of the Given is the idea that the space of reasons, the space of justifications' or warrants, extends more widely than the conceptual sphere. The extra extent of the space of reasons is supposed to allow it to incorporate non-conceptual impacts from outside the realm of thought. But we cannot really understand the relations in virtue of which a judgement is warranted except as relations within the space of concepts: *relations such as implication or probabilification, which hold between potential exercises of conceptual capacities*. The attempt to extend the scope of justificatory relations outside the conceptual sphere cannot do what it is supposed to do." (McDowell, 1996, p. 4) [emphasis mine]

McDowell too is concerned both with a relation of logic (implication) and a relation of support (probabilification), which hold between exercises of conceptual capacities, i.e., the having of

[76] It might be that when asked 'why do you think that?' one can't give a reason, but in not answering the question, one is still "…playing the game of giving and asking for reasons." (Sellars, 1997, p. 123). One is still operating within the logical space of reasons.

complex thoughts. But here it is not problematic as it is in Boyle's discussion; McDowell is explicitly talking about the warranting of judgements with complex conceptual content, which already operates in the Space of Reasons. But note that McDowell discusses the relations in virtue of which a judgement is warranted as relations in the Space of Concepts, distinct (at this stage in his dialectic) from the Space of Reasons. The talk of the Space of Concepts seems to be the space of conceptual understanding, to operate in the Space of Concepts is to have conceptual understanding of the sort the Generality Constraint is a constraint on. But like Boyle, McDowell goes further, including support relations in the space of conceptual understanding. And this is because McDowell aims to avoid the Given. To do so, he suggests the Space of Concepts and the Space of Reasons is coterminous.

Much of what McDowell says bears on what it is for a subject to *judge* that *p*, not merely exercise the concept in question, and as such (unless we take on board the suggestion that believing that *p* is not so far separated from judging that *p*), much of what McDowell says does not appear to bear on claims regarding what it is to be conceptually competent. But this, I think, is too quick. McDowell, in articulating the temptation to appeal to the Given as an ultimate warrant for a judgement says:

> "We could not begin to suppose that we understood how pointing to a bit of the Given could justify the use of a concept in judgement could, at the limit, display the judgement as knowledgeable-*unless we took this possibility of warrant to be constitutive of the concept's being what it is, and hence constitutive of its contribution to any thinkable content it figures in*, whether that of a knowledgeable, or less substantially justifiable, judgement or any other." (McDowell, 1996, p. 6) [emphasis mine]

That is, the ability to use and recognize relations of support Boyle aims to make part of the condition of having complex thought is, for McDowell a constitutive part of what it is to be the kind of concept which could figure in judgement. And the kind of concept which could figure in a judgement that things are thus and so, is the sort of concept which figures in belief. McDowell seems to give us a picture where conceptual competence demands competence not just with logical relations, but with support relations, i.e., a competence reflected by the Claiming Constraint.

4.2.4. M-expression, R-expression, and the Simple View

So far, we have the idea that an M-expression of an attitude does not entitle one to express further attitudes, while an R-expression does, and the idea of a Claiming Constraint on such expressions. If there is such a constraint on attitudes, M-expressions are not hostage to it, since there is no entitlement to further attitudes, but R-expressions are hostage to such a constraint. As suggested above, the notion of grasping support relations the Claiming Constraint appeals to can be glossed as an understanding of the reasons why an attitude is held. I suggested above that the Simple Account makes sense of the reflective moving from the assertion '*p*' to the assertion 'I believe that *p*' that R-

expression is supposed to capture. I now want to go further. Boyle's idea of R-expressing an attitude simply is an articulation of the same insight the Simple Account explains.

Further, the Claiming Constraint is intimately related to the Simple view. Recall that the claiming constraint is the following:

> For a subject S to claim that *p*, she must be entitled to a grasp of the support relations between the content of her claim and the contents of any other claims within a system of possible claims.

Implicit in this constraint is the following idea – when a subject *S* makes the assertion '*p*', for her to be entitled to a grasp of the support relations between the content of her claim and the contents of other claims, she must *understand that she is making a claim*. She must be entitled to move from the assertion '*p*' to the assertion that 'I believe that *p*'. If this were not the case, then her assertion '*p*' would be, as it were, isolated from any other assertions she would make. The Simple Account gives us an account of the entitlement to move from the assertion '*p*' to the assertion that 'I believe that *p*', and as such an account of how the Claiming Constraint can indeed be a constraint.

The Simple Account claims that the epistemic capacity exercised in making the assertion 'I believe that *p*' is nothing more than the epistemic capacity exercised in making the assertion '*p*'. If asked to provide reasons for her assertion 'I believe that *p*', the reasons the subject provides will be the very same reasons she would provide if asked to provide reasons for her assertion '*p*'. Recall that Evans' Transparency Remark is formulated in terms of answering questions; you are put in a position to answer the question 'do you believe that *p*?' by answering the question 'is it the case that *p*?' – the reasons one would give for the answer to the first question just are the reasons one would give for the answer to the second. The Claiming Constraint suggests that the speaker must be aware, in her assertion '*p*', of the reasons that support her assertion, and in this awareness, that her assertion is one amongst a network of assertions, not a logically isolated utterance. That is, her assertion is an R-Expression. An R-Expression of an attitude is an expression which is sensitive to reasons, to which the question 'why do you say that?' can be applied, and as suggested above, R-expressions are not inferentially isolated, unlike M-expressions, which can provide no basis for further expressions of belief. Sensitivity to the Claiming Constraint is, as suggested previously, a condition on R-Expression. But to be sensitive to the Claiming Constraint, the subject must, when she makes the assertion '*p*', also be in a position to make the assertion 'I believe that *p*'. As such, she understands herself as making a claim. And the Simple Account explains this.

We have not yet established 'fundamentality', however. Rather we have established that the Simple Account of Transparency implies sensitivity to reasons, which should, I think, be unsurprising given Evans' initial focus on the answering of questions in the Transparency Remark itself. To get a grip on the fundamentality of a 'deliberative' account, we need to examine the scope of such an account, and a challenge that an account of self-knowledge must engage with.

4.3. What is the Scope of the Deliberative Account, and What Does This Tell Us About Self-Knowledge?

The question which concerns Boyle is that of the scope of a deliberative account: how much of our self-knowledge is captured by a deliberative account?

Moran from the start has restricted his account to *attitudes* and excluded *sensations*, but there are other states which appear to be known immediately and with authority which are not known through making up one's mind, for example, one's appetites or one's recalcitrant attitudes (when they become available to one). Further we can (it seems) often avow our own attitudes without engaging in deliberation – e.g., when asked if we want a beer or whether we believe Edinburgh is the capital of Scotland. In these cases, one's mind is, so to speak, 'already made up':

> "The question what I believe or desire is still, of course, transparent for me to a question about what is so or what is desirable, but the relevant convictions of fact or desirability are not being formed in the present, and so it is hard to see how an appeal to agency can help to explain my present knowledge of them." (Boyle, 2009, p. 139)

Thus, it might seem that the deliberative account is quite limited in what it describes. Yet Boyle suggests that the deliberative account describes the *fundamental* type of self-knowledge, "…one that gives proper place to the immediacy of first-person awareness and the authority with which its claims are delivered." (Boyle, 2009, p. 138)

4.3.1. Fundamentality and Uniformity

Boyle points to several criticisms of a deliberative account, all of which trade on the limited scope of the account and suggest that it cannot be right based on this limited scope. Not only can the account not be fundamental, the objection suggests, it cannot even be correct. It does not provide a general explanation of self-knowledge. These criticisms are based on a central assumption regarding the business of an account of self-knowledge. The assumption is that in giving an account of any individual aspect of self-knowledge (say knowledge of attitudes, or knowledge of sensations), that account is also either an account of all self-knowledge or can be relatively trivially extended into one. The underlying thought here is that self-knowledge is in a sense homogenous, or uniform. This assumption is the *uniformity assumption*:

> "[A] satisfactory account of our self-knowledge should be fundamentally uniform, explaining all cases of "first-person authority" in the same basic way." (Boyle, 2009, p. 141)

Boyle suggests the uniformity is in terms of explaining authority (see e.g., section 1.1.1.1; in short, the idea that the subject is in a better position epistemically than an interlocutor when she asserts or avows her self-knowledge), but the assumption is trivially generalised to any interesting aspect of self-knowledge we would wish to explain[77]. If the uniformity assumption were respected, it seems that an account of self-knowledge which were truly fundamental would explain all self-knowledge within its rubric. But the deliberative account, the objection suggests, does not do this, and further, it is clear from the outset that this is not one of the explanatory goals of Moran's deliberative account. Moran proceeds from the very beginning with the assumption that the deliberative account is an account of attitudes, not sensations, i.e., that self-knowledge is not uniform. If Boyle's claim that the deliberative account is fundamental is to hold, it must be that "…"the fundamental form" does not mean "the form an account of which can serve as the model for an account of all immediate, authoritative self-knowledge."" (Boyle, 2009, p. 140) I will suggest below in section 4.3.3. that Boyle is too quick to dismiss the prospects for a unified explanation taking the insights of the deliberative account seriously, but for now we will proceed on Boyle's assumption. Boyle's suggestion is that the fundamentality of Moran's deliberative account proceeds on a different axis from a fundamentality in the explanation of self-knowledge. The deliberative account says what capacities or abilities must be in place for there to be any self-knowledge at all:

> "It is fundamental because the ability to say what one believes in the way that Moran specifies is intimately connected with the kinds of representational abilities that must be possessed by a subject who can make comprehending assertions, and a subject who lacks these sorts of abilities cannot be a self-representer, in the sense we have specified, at all". (Boyle, 2009, p. 151)

On this view of what it is for an account of self-knowledge to be 'fundamental', the deliberative account is an account of a fundamental form of self-knowledge, but it does not deliver a uniform account of self-knowledge. The deliberative account describes what must be in place for one to have self-knowledge *at all*. In that sense, in giving such an account, the deliberative account does not engage in the same project as those who aim to give a uniform account. The aim is not to give an account that says, for any given attitude, how can that attitude amount to self-knowledge. Rather the aim is to ask what must be in place for the subject to enjoy self-knowledge at all.

[77] A uniform account of self-knowledge was, for example, one of Byrne's explanatory goals in *Transparency and Self-Knowledge* (Byrne, 2018), although he did not frame it in terms of *authority* (rather in terms of privileged and peculiar access).

Boyle suggests that the explanation of these conditions for the possibility of self-knowledge means that a uniform account of self-knowledge is simply not possible; any account of self-knowledge must be sensitive to the concerns the deliberative account takes as the central topic, and as such cannot explain all self-knowledge under one umbrella. This demand of sensitivity provides a challenge to theorists of self-knowledge.

Of course, if the argument in section 4.2. of this essay holds against Boyle, the way a 'self-representer' is spelled out is distinct from the way Boyle spells it out, and the challenge that any adequate theory of self-knowledge must be sensitive to the concerns that a carefully approached explication of the deliberative account makes salient will not take the same shape as that offered by Boyle. Nevertheless, such a challenge still exists and is still a concern for a theory of self-knowledge. We shall reformulate the challenge in the next section.

4.3.2. A Challenge for Theories of Self-Knowledge

Boyle's suggestion is that self-knowledge requires the capacity for R-expression, which at least entails the capacity to represent one's mental states (and express those representations), and presents a challenge for a theory of self-knowledge:

> "Any theory of self-knowledge will confront [a] challenge […]: it must leave room for an account of what it is not just to *have* mental states but to *represent* one's own mental states." (Boyle, 2009, p. 146)

But more than this, the capacity for R-expression, as we have discussed above, is not merely the capacity to represent one's own mental states, but is the capacity to represent one's own mental states as one's own:

> "If I am right that accounting for this requires crediting the subject with a special kind of knowledge of his own deliberated attitudes, then any theory of self-knowledge must leave room for knowledge of this special kind." (Boyle, 2009, p. 146)

Crediting the subject with this special kind of knowledge is, in effect, crediting them with the competency to make utterances constrained by the Claiming Constraint.

Boyle aims to present a challenge for a satisfactory account of self-knowledge in general. Any satisfactory account of self-knowledge cannot just be an account of how a subject has particular mental states (paradigmatically beliefs), but rather it must be an account of what it is for that subject to represent those mental states. I think, however, this formulation of the challenge is a confused or mysterious in formulation: If a subject is to 'represent their mental states', what are they to represent them as? I take Boyle's point to be that the idea of representation of the mental

states in question is not merely that one's own mental states must be represented, but they must be represented *as one's own*. This, I think, is part of the lesson of sensitivity to the Claiming Constraint and R-expression discussed above. If a subject has a belief that *p*, she must not merely represent some subject as believing that *p*, she must represent herself as believing that *p*. She must, as Boyle would put it, be in a position to employ a 'form of the first person' in believing that *p*. This is a consequence of accepting the Claiming Constraint:

> For a subject S to claim that *p*, she must be entitled to a grasp of the support relations between the content of her claim and the contents of any other claims within a system of possible claims.

As suggested above in section 4.2.4, to grasp the Claiming Constraint, the subject must understand that *she* is making a claim (i.e., be in a position to employ a 'form of the first person'). Further, the suggestion above was that what it is to embrace the Claiming Constraint was to embrace the Simple Account. We can only make sense of the Claiming Constraint in light of the Simple Account.

Indeed, as we discussed above, the employment of a 'form of the first person' undermines the idea of self-representation (as understood by Boyle) as self-representation has a subject-predicate form which involves identifying knowledge of the subject. So, we need to take Boyle's (correct) insight that self-knowledge implicates an understanding of a form of the first person and reformulate the challenge without an implication of self-representation (in Boyle's pernicious sense), a reformulation which is compatible with the Simple Account. This will also let us recover the sense in which the deliberative account is fundamental for self-knowledge.

Perhaps we can reformulate Boyle's challenge like so:

> If a subject believes that *p*, she must understand (and as such be in a position to know) that she believes that *p*. Any adequate theory of self-knowledge must be sensitive to this fact, and as such, any theory of self-knowledge will confront a challenge: it must leave room for an account not just of what it is to have a belief, but what it is to understand oneself to have a belief.

Call this challenge The Challenge from the First Person. This challenge preserves the thought that there is something significant about R-expression, and that (as Boyle would put it) a form of the first person is central to self-knowledge. We should note that the challenge is given entirely in terms of beliefs, rather than mental states in general, and as such could be seen as an illicit response to the problem of uniformity for an account of self-knowledge, insofar as it is a challenge for an account of *belief* not of other states that might be self-known. I suggest that this is not so. Rather, this challenge

constrains what it is to be a believer and *inter alia* a knower *at all*. So, any account of self-*knowledge* must be sensitive to this challenge whether the account has designs on giving a unified explanation of self-knowledge or not. It is by giving an answer to this reformulation of the challenge that we can recover the sense in which a deliberative account is fundamental.

Boyle's discussion of R-expression was intended to be an answer to this challenge, but as suggested above, the idea of R-expression as being constituted in a self-representation is a non-starter, so we must answer the challenge in a different way. If instead we think of the difference between M-expression and R-expression as not being the difference between a self-representer and non-self-representer, but rather as being the difference between being able to exercise the epistemic capacity to believe and not (i.e., the difference between being a believer and not), we can move toward an answer to the challenge. The detour into self-representation Boyle takes us on is just that, a detour. We do not need a grip on the idea of a self-ascribed representation to make sense of R-expression. In fact, allowing for a self-ascription in the self-representation seems at least to imply that the ascriber has identifying knowledge of the one they are ascribing to, thus undermines the self-consciousness of the self-representation, and as such, it seems, undermines the very idea that R-expression is meant to capture:

> "It follows that a subject's use of "*A*" will *express self-consciousness* only if it bears the right sort of connection to this ability: he must understand that the person he calls "*A*" is the very person whose mind is, so to speak, his to make up." (Boyle, 2009, p. 153) [emphasis mine]

We can hold on to the idea of R-expression without self-ascription by accepting that to R-express is to exercise one's epistemic capacities in the manner suggested by the Simple account of Transparency. That is, we can meet Boyle's constraint on an acceptable account of self-knowledge by fully adopting the thought that assertions with 'I' as subject does not implicate identifying knowledge on the part of the asserting subject, that is by respecting Anscombe's point in *The First Person* (Anscombe, 1981). The Challenge in effect asks that any assertion of belief be a genuinely first-personal assertion. It is here that the idea of self-representation leads us astray. We need not posit that a speaker (thinker) self-ascribes a representation to make sense of 'A believes *p*' as a form of the first person. Anscombe's discussion tells us exactly that. What distinguishes 'I' from 'A' in Anscombe's discussion is self-consciousness:

> "The first thing to note is that our description does not include self-consciousness on the part of the people who use the name "A" as I have described it. They perhaps have no self-consciousness, though each one knows a lot about the object that he (in fact) is; and has a name, the same as everyone else has, which he uses in reports about the object that he (in fact) is." (Anscombe, 1981, pp. 24-25)

Self-consciousness is exactly what Boyle suggests the subject who can 'make up his mind' possesses that the trained parrot does not, i.e., is the difference between R-expression and M-expression. What Anscombe's discussion tells us is that whatever the self-consciousness of an assertion with 'I' as subject amounts to, it *cannot* amount to identifying knowledge of the subject of the assertion, and as such we have a minimal condition on what a self-conscious assertion can be; an assertion where the subject understands the content of the assertion but does not have identifying knowledge of that which is in the subject position of the assertion. This is where Boyle's proposal based on the self-ascription of a representation falls down, as the self-ascription requires identifying knowledge of the subject ascribed to (and as such denies that the subject of the assertion is a form of the first person, in the same way as the assertions of the 'A'-users Anscombe introduces). Further recall that I suggested that the Simple Account can accommodate the other central insight of Boyle's account, the insight that he credits Moran's deliberative account with, that in asserting *p*, a speaker is *entitled* to assert 'I believe that *p*', the insight that the digression into self-representation was initially intended to capture (with the challenge for a theory of self-knowledge arising as an attempt to capture this insight). Indeed, I suggest that the Simple Account provides a neater account of this entitlement, since no exercise of epistemic capacities beyond those exercised in the assertion *p* are required for the subject to possess the entitlement. Rather than the entitlement being a consequence of the subject being able to represent beliefs in a certain way, the entitlement is a consequence of *being a believer at all*, i.e., possessing (and articulating) their epistemic capacities. And only now are we in a position to recover the fundamentality of the deliberative view.

4.3.3. Fundamentality Recovered, Unification Explained

Even though I have suggested Boyle's model for the recovery of the fundamentality of the deliberative account of Transparency fails, I think his general point still holds, and the Simple Account delivers on the idea that the self-knowledge is fundamental insofar as it does indeed characterize a framework into which any story about self-knowledge must fit – it must answer the Challenge from the First Person. The Simple Account answers the Challenge, and in answering also tells us something about any satisfactory answer. Any satisfactory answer cannot employ a notion of self-ascription of a property to an object, since to do so would implicate identifying knowledge of the object of the ascription on the part of the ascriber, and Anscombe's point rules out such identifying knowledge on pain of not answering the Challenge at all. The Simple Account is fundamental because it tells us not merely how a subject gets some attitudes that amount to self-knowledge, but what it is to be a believer (and as such a knower) at all. The Simple Account says something about the link between self-knowledge and self-consciousness, in the same mood as Moran's deliberative account – it says that the entitlement to move from the assertion '*p*' to the assertion 'I believe that

p' by no further exercise of epistemic capacities, what Boyle and Moran might characterise as the self-consciousness of belief, is a consequence of being a believer at all. This, I suggest, where the deliberative account missteps and the Simple Account does not. The deliberative account distracts from Evans' insight by adopting the metaphor of 'making up your mind', by saying that the answer to the question 'do you believe that p?' is reached by *deliberating* over *p*. Moran and Boyle are on to the same insight as Evans, but this focus on 'deliberation' and the metaphor of 'making up your mind' hides the great insight that the Simple Account brings this out. We can recover this insight if we realise what the deliberative account should say is that talk of 'making up your mind or of taking a deliberative stance over one's belief is a metaphorical way of formulating the idea that one makes the assertion 'I believe that *p*' by no further exercise of an epistemic capacity than those exercised in making the assertion '*p*'. Thus, we can see why the self-knowledge described by the Simple Account of Transparency is fundamental. No account of self-knowledge can proceed without the recognition that the subject who enjoys such knowledge is a believer, and the Simple Account puts a constraint on what it is to be a believer at all. This sort of 'fundamentality' is exactly the sort Boyle argues for, which motivates his denial of the Uniformity Assumption.

As Boyle puts it:

> "An account of self-knowledge which accepts the Uniformity Assumption must either rule out the kind of self-knowledge Moran describes, or else maintain that all of our self-knowledge is of this kind. Everyone agrees that the latter option is untenable." (Boyle, 2009, p. 156)

That is, if an interlocutor wishes to avail themselves of the scope objection to a deliberative account – that a deliberative account cannot deliver an account of all self-knowledge, and as such, because the Uniformity Assumption holds, a deliberative account fails as an account of self-knowledge and should be scrapped – they, the interlocutor, are put in an untenable position.

The Simple Account does not as such give an account of all attitudes or states that might fall under the rubric of self-knowledge, nor does it aim to. The aim is to provide an account of belief, and in doing so say something about the nature of being a believer at all. Other states such as desires, or sensations, or intentions, are simply not within the scope of the Simple Account itself[78]. If the Simple Account tells us something about the fundamental nature of being a believer, then it seems the supporter of the Uniformity Assumption cannot simply rule it out on the grounds of the scope of the Simple Account's explanatory power.

[78] Although, as discussed below, what the Simple Account does do is tell us something about the *form* of an account of self-knowledge.

We are in effect presented with a crude dilemma whose horns are 'accept a deliberative account as fundamental and concede that it explains all self-knowledge, holding on to the Uniformity Assumption' or 'accept a deliberative account as fundamental and that it does not explain all self-knowledge, thus rejecting the Uniformity Assumption'. The idea is that the first horn is a non-starter. Even the supporter of the deliberative account in the offing concedes that it does not explain all self-knowledge and nor is it intended to, and so we must reject the Uniformity Assumption and concede that different aspects of our self-knowledge admit to different explanations. Nevertheless, the thought goes, the deliberative account tells us something important about being a self-knower *at all*, so even if 'belief' is a small slice of the self-known states of a subject, it is a fundamental slice, and as such the explanation offered by the deliberative account is still theoretically interesting.

This crude dilemma is, I suggest, just that, crude. There is a third way, between the horns of the dilemma. The insight of the Simple Account is that the baggage of agency and 'making up your mind' is a red herring. We can make sense of the 'deliberative account' without deliberation (as it were) – this is in effect the business of Chapter three of this work, where the Simple Account was developed. But this doesn't mean that we should deny that Boyle and Moran are on to something regarding the fundamentality of the account. Indeed, I hope I have shown that such a denial would be deeply misplaced. Rather, I suggest that Boyle does show that there's something fundamental about 'deliberation', which I have articulated above, but that claim to fundamentality does not mean we should accept the horn of the dilemma which denies unification.

We can hold on to the unification of self-knowledge by abstracting away from the details of the deliberative account and instead examining the *form* of an account which abstracts away from the details of a deliberative' account of (self-knowledge of) belief. Recall that the great insight from Evans is that to answer the question 'do you believe that *p*?', I need exercise no further capacity than those exercised in answering the question 'is it the case that *p*?'. In this case, the case of the transparency of belief, the capacity in question is an epistemic capacity, the capacity to believe. In exercising the capacity to answer a question regarding how things are, I am, by exercising no further capacities, in a position to claim a piece of self-knowledge, knowledge of what I believe. There is a suggestion of form here that can perhaps be extended into a general claim about the form an account of self-knowledge in a domain should take (and as such a claim about the form of self-knowledge in general). Take the account of belief developed in this work; Evans' insight is that self-knowledge of belief is not achieved by an exercise of a separable epistemic capacity from the capacity which is exercised in the formation of the belief which is the topic of the self-knowledge in question. Exercise of the capacity to believe and self-knowledge of the belief one forms by that exercise are in this sense one. This unity between that which is the topic of the self-knowledge and

that which is self-known provides us with a form of explanation for self-knowledge in other domains. The exercise of the capacity that constitutes the [formation/possession] of the state or attitude in question also constitutes knowledge of the [formation/possession] of that state or attitude. The general form of an account of self-knowledge would, on this view, be the following:

> To have self-knowledge of her being *F*, the subject need do nothing more than exercise the capacity the exercise of which is her being *F*.

This is Evans' insight regarding belief: self-knowledge of belief is had by doing nothing more than exercise the capacity to believe. In this way we can preserve a unified account of self-knowledge while holding on to Boyle's idea of fundamentality. The uniformity is not in the capacities implicated in the explanation of self-knowledge, because the cognitive capacity exercised in the formation of belief need not be related to the cognitive capacity exercised in the formation of e.g., intention. But nevertheless, there is still uniformity, uniformity of the form of explanation, suggested by the general form of an account of self-knowledge above. To use the example of intention, an account of the self-knowledge of intentions would follow the general form:

> To have self-knowledge of her intention to do A, the subject need do nothing more than exercise the capacity the exercise of which is her intending to do A.

Of course, this is no account of the subject's self-knowledge of her intentions without an account of the capacities exercised in the having of intentions, and I have delivered no such thing. To do so would be beyond the scope of this work, as giving an account of such capacities would be to give an account of practical reason in general. Nevertheless, this is no mark against the claim that what Evans' insight gives us is a structural understanding on self-knowledge, which meets a uniformity constraint and gives a general understanding of what it is for something to be self-knowledge. Note also that although the uniform explanation on offer is in terms of the *form* an account of self-knowledge takes, the explanation on offer is still in Byrne's terms an *economical* one; Evans' insight might be put this way; there is no special capacity for self-knowledge. Insofar as an account of self-knowledge posits epistemic capacities at all, it does not posit any epistemic capacities that are specific to self-knowledge. This is at least close to how Byrne characterises economy:

> "Let us say that a theory of self-knowledge is *economical* just in case it explains self-knowledge [without positing any] epistemic capacities and abilities [other than those that are] needed for knowledge of other subject matters; otherwise it is *extravagant*." (Byrne, 2018, p. 14)[79]

[79] There are some slight changes to Byrne's formulation, noted in square brackets. I do this to clean up the formulation in line with what I have said in this chapter. I suggest nothing important is lost in making these small changes.

Byrne's characterisation is in terms of *epistemic* capacities, but I see no reason to restrict economy in this way, given the focus on form above. We can perhaps understand economy like this: the form of an explanation of self-knowledge is economical just in case it posits no special capacities or abilities to explain self-knowledge in a domain beyond those capacities implicated in the explanation of the domain itself.

Of course, an interlocutor who wants unification at the level of the epistemic capacities implicated in the explanation of self-knowledge will cry foul at the move to the abstract level of form. But such an interlocutor is left on the horns of the dilemma Boyle sets up, between the unification of self-knowledge and the fundamentality of the deliberative account. Further, the desire for unity at the level of epistemic capacities asks for more from a uniformity assumption than Boyle suggests. Recall that Boyle's uniformity assumption is this:

> "[A] satisfactory account of our self-knowledge should be fundamentally uniform, explaining all cases of "first-person authority" in the same basic way." (Boyle, 2009, p. 141)

The suggestion at the level of form meets this constraint. All self-knowledge is explained in the same basic way, but not by appeal to the same capacities. The abstraction to the level of form allows us to navigate the dilemma Boyle sets, and still and retain a unified account of self-knowledge, and the fundamentality of a deliberative account.

4.4. The Objection from Scope and Sensations

The response suggested for the Objection from Scope is to provide a general formulation of a Transparency account that extends the basic insight of the Simple Account from belief to further domains of self-knowledge. The generalisation of the Simple Account took the following form:

> To have self-knowledge of being *F*, the subject need do nothing more than exercise the capacity the exercise of which is her being *F*.

Recall that the central insight of the Simple Account is that one gets oneself into a position to have knowledge that one believes *p* by doing nothing more than exercising the epistemic capacity to believe *p*. Self-knowledge is had 'for free' as it were. The generalisation takes this insight and applies it to domains of self-knowledge outside of belief[80].

Above, I motivated this generalisation by suggesting that self-knowledge of intention provided a plausible extension, although as noted there providing a fully worked out account of self-knowledge of intention on the model suggested is beyond the scope of this thesis, as it would amount to a worked-out account of practical reason. The suggestion is rather that this general formulation can be

[80] Indeed, the suggestion is that this general form is the form an account of self-knowledge in general takes.

the basis of a research project focused on extending the Simple Account to self-knowledge in general.

4.4.1. Conceptualism

One worry, though, is that the resolution of the Objection from Scope is contingent on this extensibility of the general account, and although the account seems to extend obviously into some domains of self-knowledge, like intention, other domains may prove less congenial. We might think that the general extension of the Simple Account seems natural in those domains of self-knowledge where the object of self-knowledge is *conceptual* and is perhaps more challenging in those domains where the object of self-knowledge is *non-conceptual* (for example, Moran's distinction between sensations and attitudes could be understood as a division between those objects of self-knowledge that are conceptual and those that are not obviously so[81]). Centrally, we might worry that extension to the domain of self-knowledge of sensation presents a challenge to the Simple Account in the following way: If the Simple Account endorses a *conceptualist* account in some domain of self-knowledge (e.g., self-knowledge of belief), then extending the account across domains using the general formulation above endorses conceptualism about the objects of self-knowledge in general. If this challenge holds, then it makes the response of the Simple Account to the Objection from Scope contingent on an otherwise controversial philosophical position that has not been independently motivated – that of thoroughgoing conceptualism regarding the objects of self-knowledge.

4.4.2. The Challenge from Conceptualism

To understand this challenge, we should understand what is meant by and entailed by conceptualism in this context. Following McDowell, we can understand the conceptualist position regarding *experience* as the idea that experience is conceptually structured throughout – in seeing (i.e., experiencing) a blue cup (say) one is not delivered non-conceptual content by perception which is then conceptualised as a blue cup. Rather, perception (experience) is an actualization of conceptual capacities. McDowell puts this in terms of *receptivity* and *spontaneity*, where receptivity is the capacity of a subject to sense – her receptiveness to the world, and spontaneity is her capacity for conceptual thought. The central conceptualist idea is that in empirical knowledge receptivity and spontaneity are (as McDowell puts it) not notionally separable – sensing *is* thinking, in a slogan. There is no controversy that e.g., beliefs and judgements bear conceptual content, and this is where the Simple account begins, but it is not a settled matter as to whether sensations bear conceptual

[81] Moran (2001), for example, is explicit that his deliberative account is restricted to attitudes and excludes sensations

content, and a tacit commitment to this would be a substantial philosophical claim which has not been independently argued for.

This provides a challenge for the Simple Account as the picture the Simple Account can suggest for sensations is that some capacity is exercised in having the sensation *S* (the capacity to have pain, for example), and by nothing more than the exercise of that capacity, the subject is in a position to know she is in/has *S*. The conceptualist objection is that this picture entails an unwanted conceptualism, because the move from having *S* to knowing one has *S* entails that *S* is a state with conceptual content.

The response on behalf of the Simple Account is to say that the generalisation of the Simple Account says that the exercise of the capacity the exercise of which is having a sensation *S* puts a subject *who is suitably conceptually equipped* in a position to know she has *S*, i.e. so long as she has the concept of being *S,* the exercise of the capacity to have *S* puts the subject in a position to know she has *S*. This does not entail that *S* itself is a state with conceptual content, or that the exercise of the capacity exercised in having *S* is a conceptual act (an exercise of a conceptual capacity). The Simple Account is *neutral* with regards to Conceptualism. A subject who does not have the appropriate concepts, when they exercise the capacity to have *S is not in a position to know they are in S* – they lack the appropriate conceptual resources for self-knowledge.
The central point is that in a conceptually equipped subject, to (be in a position to) know one is in a state *S*, one need do nothing more than exercise the capacity whose exercise is the having of *S*, whether or not the content of *S* is conceptual. Self-knowledge (as it were) comes along for free in appropriately conceptually equipped subjects. Although the subtlety regarding appropriately conceptually equipped subjects is not explicit in the general formulation of the Simple Account, I suggest it follows naturally from what has already been said. This is no revision of the account, but rather a making explicit of something already assumed in the theory – self-knowledge is available to appropriately conceptually equipped subjects, and it is no epistemic achievement *in general*.

4.5. Concluding Remarks to Chapter 4

In this chapter the central focus was on deflecting the Objection from Scope. In doing this, we developed the idea that a 'deliberative' account of belief which develops from the Transparency Remark is *fundamental*. But unlike Boyle, we reject the dilemma between this fundamentality and the unification of self-knowledge. We do this by considering the *form* that an account which is developed from the Transparency Remark (properly understood) takes. In getting to this point we preserved the insights Boyle took from Moran's deliberative account while at the same time

rejecting the mistaken application of self-representation. These insights allowed the explication of a constraint on belief and so, given the Simple Account, self-knowledge of belief, the Claiming Constraint, which helps to make explicit the sensitivity to reasons implicit in Evans' formulation of the Transparency Remark in terms of answering questions. With the Objection from Scope successfully deflected, we are now in a position to see that the Simple Account of Transparency can engage with and dispose of the central objections to Transparency accounts, and although the Simple Account only explains self-knowledge of belief, it reveals the general form of an account of self-knowledge and opens the way for further research making good on this promise. So far, however, we have only engaged with Rationalist accounts of the Transparency Remark. In the next chapter, we will examine the central Inferentialist development of the Transparency Remark as the main competitor to the Simple Account, Alex Byrne's account developed in *Transparency and Self-Knowledge* (Byrne, 2018).

5. Byrne's *Transparency and Self-Knowledge*

Alex Byrne, in *Transparency and Self-Knowledge* (Byrne, 2018) develops an account of self-knowledge that seeks to build on Evans' basic insight that the assertion 'I believe that p' is in some sense 'transparent' to the assertion '*p*'. And further, like Evans, Byrne suggests that the 'transparent process' by which one gains self-knowledge of one's belief that *p* is accomplished by one's attending to evidence which bears on *p*[82]:

> "If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting in to operation whatever procedure I have for answering the question whether *p*." (Evans, 1982, p. 225)

This itself does not suggest what the procedure in question is. Byrne understands this 'transparent procedure' as an *inferential transition* between contents. Byrne suggests that the transparency of self-knowledge is best understood as an inference from the fact that *it is raining* to the conclusion that one believes that *it is raining*:

> "But what is it, exactly, to discover that one believes that it's raining "by considering" the weather? If one discovers that there are mice in the kitchen by considering the nibbled cheese, one has inferred that there are mice in the kitchen from premises about the nibbled cheese. So an obvious and natural way of cashing out the transparency of belief is that one's knowledge that one believes that it's raining is the result of an inference from premises about the weather." (Byrne, 2018, p. 74)

In this chapter I will aim to understand whether Byrne's account is an account of Evans' Transparency Remark, and whether it can be understood as an account of self-knowledge for what we might consider the central case; that of each subject's knowledge of her own beliefs[83]. In doing this, I aim to suggest a possible lacuna in Byrne's account which must be examined and explained for his account to be satisfactory by his own lights, and ultimately, I aim to show that although Byrne's account might in the end be an Inferentialist account of self-knowledge, it is no development of the Transparency Remark, but is rather the denial of Evans' position.

[82] In the rest of the chapter, whenever possible, we will replace the empty variable *p* in the statements of the transparent inference which are not direct quotations with the statement '*it is raining*'. The statement will be italicised throughout. The aim is not to undermine the generality of the Transparency point, but merely for the purpose of expositional and argumentative clarity.

[83] Byrne's method throughout the positive account in *Transparency and Self-Knowledge* is to build on or extend the account he develops for belief. As such, his account of belief is central to his account of Transparent self-knowledge, it is the foundation his account is built on. His account stands or falls with the account of belief.

5.1. Byrne's Explanatory Goals

Byrne aims to give an account of self-knowledge which respects Evans' Transparency Remark but avoids its puzzling aspects[84]. In doing this, Byrne aims to explain what he suggests are two central aspects of self-knowledge: *privileged access* and *peculiar access* and aims to do so by providing what he describes as an *economical detectivist* account. Here we will engage in some ground-clearing regarding what Byrne takes himself to be doing.

5.1.1. Privileged Access

Privileged access is the thesis that

> "…beliefs about one's mental states acquired through the usual route are more likely to amount to knowledge than beliefs about others' mental states (and, more generally, corresponding beliefs about one's environment)." (Byrne, 2018, p. 5)

It is at least prima facie plausible that we know our own minds better than we know the minds of others or facts about the world, so an explanatory goal of a sufficient account of self-knowledge is that it explains such privileged access[85].

5.1.2. Peculiar Access

Peculiar access is the thesis that

> "…one can come to know about one's mental life "in a way that is available to no one else.""
> (Byrne, 2018, p. 8)

The general idea is that each one of us knows our own mental life not through behavioural or third personal cues or evidence, but via a method that each thinker can only apply to their own mental life. Note that this does not rule out behavioural or other third personal evidence as evidence of one's mental life, but rather complements and adds on to it. Again, by Byrne's lights, a satisfactory account of self-knowledge must account for peculiar access[86].

5.2. The Shape of An Account of Self-Knowledge

Byrne suggests that there are four axes which determine the shape of an account of self-knowledge. These axes are *economy-extravagance*, *detectivist-non-detectivist*, *inferential-non inferential* and *unified-non-unified*.

---

[84] See the standard objections to Transparency in section 1.4. Byrne is concerned with his elucidation of the Puzzle of Transparency in particular, but there is at least tacit engagement with the other standard objections. See section 5.7.

[85] Recall that in chapter one we divided Privileged Access into *Authority* and *Groundlessness*. Byrne here seems to be only concerned with Authority, and this is not unexpected given that Groundlessness appears to exclude an inferential account of Transparency.

[86] Byrne's characterisation of Peculiar Access shares some features with Groundlessness but is not the same idea. Peculiar access says something about the method by which a subject comes to self-knowledge, but does not say anything about the grounds of that knowledge (except where the method rules out some grounds).

5.2.1. Economy-Extravagance

The economy-extravagance axis concerns the explanation of how one has self-knowledge. An economical account is an account which does not posit a special or unique capacity by which one has *self*-knowledge, instead relying on capacities for knowledge in general, whereas an extravagant account does posit such a capacity:

> "Let us say that a theory of self-knowledge is *economical* just in case it explains self-knowledge solely in terms of epistemic capacities and abilities that are needed for knowledge of other subject matters; otherwise it is *extravagant*." (Byrne, 2018, p. 14)

5.2.2. Detectivist-Non-Detectivist

Detectivist accounts treat self-knowledge in a manner similar to empirical knowledge:

> "First, causal mechanisms play an essential role in the acquisition of such knowledge, linking one's knowledge with its subject matter. Second, the known facts are not dependent in any exciting sense on the availability of methods for detecting them, or on the knowledge of them—in particular, they could have obtained forever unknown." (Byrne, 2018, p. 15)

Detectivist accounts say that there is nothing particularly strange about self-knowledge when compared to empirical knowledge. One's knowledge that e.g., one believes that *it is raining* is causally linked with one's belief that *it is raining*. Further, there is no special connection between the fact known in self-knowledge and the method by which one comes to know that fact.[87] A non-detectivist account does not meet either or both of these two conditions.

5.2.3. Inferential-Non-Inferential

An inferential account suggests self-knowledge is the result of an *inference*, understood as theoretical reasoning. Inference on this view is understood to involve (minimally) causal transitions between belief states[88]: "…if one reasons from P to Q (or, equivalently, infers P from Q), one's belief in P causes one's belief in Q." (Byrne, 2018, p. 15). Byrne suggests Evans' transparent procedure is naturally understood as an inference. A non-inferential account of self-knowledge does not treat self-knowledge as the result of an inferential transition.

5.2.4. Unified-Non-Unified

A unified account of self-knowledge explains all self-knowledge the subject might possess in one explanatory swoop: "For any mental state M, the account of how I know I am in M is broadly the same…" (Byrne, 2018, p. 16)[89]. A non-unified account offers individualised accounts for different

[87] For example, a constitutive connection between believing that *it is raining* and knowing that one believes that *it is raining*.

[88] It is unclear how much more than causal transitions between belief states would be needed for an account to be inferential, but Byrne seems to want to implicate reasoning as well as causation.

[89] Note that Byrne restricts the explanatory domain of an account of self-knowledge to mental states. An account of self-knowledge is an account of a subject's knowledge of her mental states. The general formulation provided by the Simple Account in chapter four is not restricted to mental states in this way.

mental states, e.g., the explanation of how we know our own beliefs differs from the explanation of how we know our own sensations which differs from how we know our own desires.

Byrne aims to offer an *economical, detectivist, unified, inferential* account of self-knowledge.

5.3. Transparency as an Inference

Byrne characterises the transition from the fact that *it is raining* to the conclusion that one believes that *it is raining* as an inference from *world to mind* (Byrne, 2018, p. 75).

Byrne suggests that a natural way of coming to understand the transparent procedure is by unpacking what it would be to believe some conclusion by 'attending to some phenomena' and drawing a conclusion from this attention. One concludes there are mice in the kitchen by attending to ("considering") the nibbled cheese and inferring that there are mice from the premiss that the cheese is nibbled, plus some background beliefs. Likewise, in Evans' example, one comes to a belief regarding one's attitude toward the possibility of a third world war by attending to (considering) premises regarding the possibility of a third world war. It is in this way that Byrne suggests the transparent procedure is best understood as an inference, from premises about the world to a conclusion regarding one's beliefs: "…one's knowledge that one believes that it's raining is the result of an inference from premises about the weather." (Byrne, 2018, p. 74)

To explain the structure of this transparent inference, Byrne adopts terminology drawn from Gallios (1996). Gallois draws on Moore's Paradox to suggest a schema for inference regarding the beliefs of a subject. The Moorean paradox suggests that it is absurd[90] to assert a statement of the form '*it is raining* but I don't believe that *it is raining*.' Gallois uses this absurdity to suggest the following: "…there is something amiss with my saying, or thinking, that p, but also saying, or thinking that I do not believe p." (Gallois, 1996, p. 46). From this recognition of the inappropriateness of such an utterance or thought, he suggests what he calls the *doxastic schema*:

p

I believe that p

The doxastic schema suggests that if the subject accepts that *it is raining*, then the subject ought to accept that she believes that *it is raining*. This is the schema that Byrne uses to make sense of the inference from world to mind he understands the transparent transition to take.

[90] Gallois suggests that "It is notoriously difficult to say [what sort of absurdity is involved]." (Gallois, 1996, pp. 45-56) (clarification of context mine). We will not focus on the nature of the absurdity here. Rather, we will take the prima facie absurdity enough to motivate what follows.

5.3.1. How Does the Inference Explain Self-Knowledge?

An inference in the form of the Doxastic Schema is, as Byrne recognises, a bad inference. There is nothing about its being the case that *it is raining* that mandates that one should infer *anything whatsoever* about what one believes. Nothing settles it that the fact that *it is raining* means that the subject *S* believes that *it is raining*, or that *S* ought to believe that *it is raining*.

How, then, are we to understand the doxastic schema as securing self-knowledge? This question is how Byrne understands the Puzzle of Transparency. The puzzle surrounding the inference from world to mind can, according to Byrne, be made manifest in three different ways, depending on where one places the emphasis in thinking about the puzzle[91]:

1) Reliability: How can reasoning in the pattern of the Doxastic Schema lead to reliably true beliefs? The premiss that *it is raining* could easily obtain without my believing it, so an argument in the form of the Doxastic Schema cannot prima facie be a reliable way to reach the truth.
2) Evidence: that *it is raining* is insufficient evidence for the conclusion that I believe that *it is raining* but according to the schema, it is my total relevant evidence. So how can knowledge be based on this inadequate evidence?
3) Reasoning through a false step: it could conceivably be false that *it is raining*, yet by following the transparent procedure I come to know that I believe that *it is raining* despite it not raining – how can it be that reasoning to a conclusion via the transparent procedure can amount to knowledge of that conclusion when, even if the conclusion is true, that reasoning proceeds through a false step? As Byrne puts it "…typically when one reaches a true conclusion by reasoning through a false step, one does not know the conclusion" (Byrne, 2018, p. 107)

Byrne puts the solution that he suggests to the Puzzle of Transparency (in each of the three versions above) by recasting the puzzle not in terms of the making of an inference from world to mind, but rather as the subject following an *epistemic rule*.

5.3.2. Inference and Epistemic Rules

Recast this way, Byrne understands the process of obtaining self-knowledge as a matter of following a special sort of epistemic rule, captured in the following formulation:

[91] It is important to note that the emphasised versions of the puzzle below make no mention of an inference but instead talk of 'reasoning'. Byrne equates these two: The psychological process of reasoning (or inferring) can extend one's knowledge." (Byrne, 2018, p. 100). I will not quibble with this equation except to note its presence – nothing in Byrne's argument or my response to it turns on concerns specific to inference but not to reasoning simpliciter.

> "BEL: If *p*, believe that you believe that *p*" (Byrne, 2018, p. 102)

BEL is an instance of a general framework, the framework of epistemic rules. To understand BEL, we need to understand this epistemic rule framework as a general framework then apply the understanding to the specific case of BEL.

Byrne suggests an epistemic rule is a conditional with the following form:

> (R) "If conditions C obtain, believe that *p*" (Byrne, 2018, p. 101)

Where following such a rule is captured by a procedure like so:

> "…S follows the rule R ('If conditions C obtain, believe that *p*') on a particular occasion iff on that occasion:
>
> (i) S Believes that *p* because she recognises conditions C obtain
>
> which implies
>
> (ii) S recognizes (hence knows) that conditions C obtain
> (iii) Conditions C obtain
> (iv) S believes that *p*" (Byrne, 2018, pp. 101-102)

Substituting BEL for the generic epistemic rule in the schema gives the following:

> S follows the rule BEL ('If *p*, believe that you believe that *p*') on a particular occasion iff on that occasion:
>
> (i) S Believes that she believes that *p* because she recognises that *p*
>
> which implies
>
> (ii) S recognizes (hence knows) that *p*
> (iii) *p* obtains
> (iv) S believes that she believes that *p*[92]

Condition (i) in the schema above suggests that S forms the belief in question *because* she recognizes C-conditions obtain, and the because here is understood to "…mark the kind of reason-giving causal connection that is often discussed under the rubric of 'the basing relation'." (Byrne, 2018, p. 101) We should note here that Byrne brings in talk of a 'basing relation' without further elaboration. I take it that he understands a basing relation to be such that the basis of my belief is a reason for and the cause of my belief (i.e., there is a reason-giving causal connection between the basis and the belief). As such, talk of 'the basing relation' may profitably be replaced by the idea of a reason-giving causal connection.

[92] In substituting BEL into the generic rubric for an epistemic rule, we have not replaced the variable *p* with a concrete example. Nevertheless, we can easily substitute our concrete example *it is raining* and see how S, upon following BEL reaches the belief that she believes that *it is raining*. We will continue to substitute the variable for our concrete example in what follows.

5.3.3. Reason Giving Causal Connections

In the understanding of the schema above, the 'because' in (i) marks a basing relation understood as a 'reason giving causal connection' between what is reported on the left-hand side of the 'because' and what is reported on the right. We might ask, however, why Byrne insists upon the 'because' in the schema being a *reason-giving* causal connection. The story provided by BEL would, it seems, work quite well as a purely causal-reliabilist story of belief-formation, without need to appeal to reasons or rationalizations[93]. That Byrne understands the transparent process as a causal-rational process appears important to his account, so the talk of reasons must do philosophical work in his account.

We can see this causal connection as a rationalizing one by examining Byrne's example of an epistemic rule:

> "DOORBELL: If the doorbell rings, believe someone is at the door" (Byrne, 2018, p. 101)

Recall the general form of an epistemic rule:

> (R) "If conditions C obtain, believe that *p*" (Byrne, 2018, p. 101)

Conditions C of the rule DOORBELL is the ringing of the doorbell. The follower of DOORBELL (Mrs Hudson in Byrne's example) believes there is someone at the door *because* she recognizes that the doorbell is ringing, and this 'because' marks a reason-giving causal connection between the recognition of the ringing of the doorbell and the belief that there is someone at the door. Working through the schema for epistemic rules above:

(i) Mrs Hudson believes there is someone at the door because she recognises that the doorbell is ringing

Which implies

(ii) Mrs Hudson recognizes (hence knows) the doorbell is ringing
(iii) The doorbell is ringing
(iv) Mrs Hudson believes there is someone at the door.

We will return to the role a reason-giving causal connection might play in Byrne's dialectic in section 5.6. once we have a fuller picture of Byrne's account in place.

[93] This would, it seems, rule out the account relying on *inference*, but the account would stand with inference replaced by 'reliable causal transition'. See section 5.6. of this chapter a discussion of this.

5.3.4. Understanding BEL as an Epistemic Rule

Byrne suggests that BEL is both a 'schematic' rule and a 'neutral' rule. We will discuss 'neutrality' in more detail below, but "[o]ne *follows* a schematic rule just in case one follows a rule that is an instance of the schematic rule; a schematic rule is *good* to the extent that its instances are."(Byrne, 2018, p. 102). BEL is a schematic rule – it gives a schema by which one can gain higher order knowledge. BEL says that

> "If *p*, believe that you believe that *p*" (Byrne, 2018, p. 102)

So, BEL is schematic insofar as we can fill in any appropriate *p* and following the rule outputs a higher order belief (the belief that one believes that *p*).

With this in place, to understand the epistemic rule BEL, and how BEL can lead to knowledge, we need to understand Byrne's use of the following:

(a) The 'neutrality' of the epistemic rule.
(b) The nature of C-conditions in BEL.
(c) The nature of the transition.
(d) How such a transition can constitute knowledge at all.

5.3.4.1. Neutrality

Byrne describes BEL as a 'neutral' rule, where neutral is understood as the following:

> "If the antecedent conditions C of an epistemic rule R do not require evidence about the rule-follower's mental states in order to be known, R is *neutral*. A schematic rule is neutral just in case some of its instances are." (Byrne, 2018, p. 102)

Since a neutral rule is not specified in terms of the follower's mental states, "…the claim that S can follow a neutral rule does not presuppose that S has the capacity for self-knowledge." (Byrne, 2018, p. 102) BEL must be a neutral rule because Byrne's account aims to explain self-knowledge, and BEL (and the broader epistemic rule account) is the explanation of how we have self-knowledge. If the antecedent C-conditions of BEL included the rule-follower's mental states, then when the rule follower engages in the following of BEL, condition (ii) would explicitly involve recognition (and hence knowledge) of the rule follower's mental states. Self-knowledge would already be present in the schema intended to explain self-knowledge, and Byrne's account would be viciously circular. Note, however, that the neutrality of BEL is not such that it generates a sceptical worry:

> "Self-knowledge is our topic, not scepticism: knowledge of one's environment (including others' actions and mental states and *reasoning (specifically, rule-following of the kind just sketched)* can be taken for granted. So in the present text, it is not in dispute that we follow neutral rules, including neutral rules with mentalistic fillings for '*p*', like 'If S has a rash, believe that S feels itchy'…" (Byrne, 2018, p. 102)(emphasis mine).

The emphasis in the above quote suggests that Byrne believes that even though the epistemic rule one follows when making the transparent inference must be neutral (i.e., the C-conditions of the rule must not themselves be mental states), the process involved in the following of the rule is still a reasoning process, and this reasoning process, we must assume, does not itself presuppose a sort of self-knowledge. This, it seems, is what Byrne means by the claim that the process of following BEL (of making the transparent inference) is not *critical reasoning* in Burge's sense:

> "Critical reasoning is reasoning that involves an ability to recognize and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. As a critical reasoner, one not only reasons. One recognizes reasons as reasons." (Burge, 1996, p. 98).

5.3.4.2. C-conditions in BEL

As suggested in our discussion of epistemic rules in general above, the C-condition of an epistemic rule is the antecedent of the conditional expressed in the rule. So for BEL, the C-condition is the antecedent of the conditional "if *p*, believe that you believe that *p*"(Byrne, 2018, p. 102), so the C-conditions for BEL are *p*. But *p* here is merely an empty variable. What we are interested in is the sort of thing that can be substituted for *p*. The motivation of this sort of account is Evans' original insight, with which we are now very familiar:

> "If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting in to operation whatever procedure I have for answering the question whether *p*." (Evans, 1982, p. 225)

Given that one attends to how things are with the world to gain knowledge of what one believes, it seems that the *p* in BEL should be whatever it is one would believe when asked 'is it the case that *p*?'. To take Evans' example and use the schematic rule BEL, we would have:

> $BEL_{World\ War}$ If there is going to be a third world war, believe that you believe there is going to be a third world war.

Again, working through the schema Byrne supplies for a generic epistemic rule, we have

> *S* follows $BEL_{World\ War}$ ('If there is going to be a third world war, believe that you believe there is going to be a third world war') on a particular occasion iff on that occasion:
>
> (i) *S* Believes that (she believes that there is going to be a third world war) because she recognises that there is going to be a third world war
>
> which implies
>
> (ii) *S* recognizes (hence knows) that there is going to be a third world war

(iii) There is going to be a third world war

(iv) *S* believes that she believes there is going to be a third world war.

Prima facie following this rule seems like a bad rule, because the initial stage of the transition is a recognition of a state of affairs in the world and the output of the transition is a higher order belief, with no intervening stage where the subject *S* believes there is going to be a third world war (i.e., the higher order belief seems to be held without the first order belief). But, of course, *S* follows $BEL_{World\ War}$ if and only if *S recognizes (hence knows) that there is going to be a third world war*. The very following of $BEL_{World\ War}$ guarantees it that *S* believes that there will be a third world war, so there is still first order knowledge, and $BEL_{World\ War}$ is still neutral because the antecedent of the rule is still not specified in terms of mental states of *S*, and what goes for this instance goes for all other good instances of BEL.

5.3.4.3. The Stages of the Transition

We have already specified some demands on the stages of the transition BEL describes. We have suggested that the antecedent of the rule BEL forms the C-conditions in the schematic form of the rule, and Byrne suggests that the C-conditions themselves are the *antecedent conditions* of the rule. We understand Byrne here to mean *not merely* that the C-conditions are the antecedent of the rule, but rather the C-conditions are the conditions under which the following of the rule becomes possible. The initial stage of the transition is specified by the C-conditions, so the initial stage is something which is, when *S* recognizes it, causal-reason giving. If the antecedent condition did not obtain, it would be impossible to follow the rule. Consider again $BEL_{World\ War}$:

> If there is going to be a third world war, believe that you believe there is going to be a third world war.

If the antecedent condition were not met, if there were not going to be a third world war, *S* could not follow $BEL_{World\ War}$

This further makes sense of Byrne's account as attempting to build upon Evans' insight that when one is asked whether one believes that *p* one comes to know what one believes by attending to the question 'is it the case that *p*?'. If the epistemic transition licenced by BEL is the procedure by which each one of us comes to know our own beliefs, then *p* being both the antecedent condition of the following of the rule, and the antecedent of the rule itself puts the fact of *p* at the centre of the formation of our higher order beliefs and therefore any account of higher order knowledge.

The state which forms the consequent of the transition licensed by BEL is a higher order belief – *S*'s belief that she believes *p*. *S*'s belief that *p* takes no place in the epistemic rule which licenses the

transition between states. Rather it appears in the schema which explains the following of the rule. *S* makes a transition from *p* to a belief that she believes *p* by following the schema, and the schema ensures that *S* both recognises and believes that *p*, and that *p* actually obtains. Byrne's insistence on recognition (a factive notion) in the schema for epistemic rules presents us with a possible lacuna. It seems that specifying the epistemic rule requires that one does not merely believe that the C-condition obtains, but that one recognizes that it does. Why does Byrne insist on recognition?

Perhaps the insistence can be explained by recalling that the transition licensed by BEL is a transition between a fact about the world (*it is raining*) and a doxastic state of an agent (*S*'s belief that she believes *it is raining*). We might wonder how this makes any sense whatsoever – how does the subject make a transition from a non-doxastic fact about the world to a doxastic state? The schema allows us to make sense of this transition. The key step in the schema to make sense of this is step (ii). The agent must *recognize* the fact about the world which forms the antecedent of the rule. The subject's recognitional capacity for states of affairs in the world is a background capacity for the possibility of the transition at all. Byrne of course understands this, and makes clear that our topic is "…not scepticism: knowledge of one's environment […] can be taken for granted…" (Byrne, 2018, p. 102) We must understand the subject as possessing certain capacities to make sense of their being able to follow BEL at all.

A more pressing question is how following a rule like BEL can result in knowledge at all.

5.3.4.4. BEL and Knowledge.

Byrne thus far has suggested a procedure a knower might go through which secures that they have second-order beliefs (beliefs about their beliefs) regarding some *p*'s. But so far, for all Byrne has said, following BEL only guarantees a specific sort of belief, a second-order one. Knowledge has not yet been secured. Transparency presents us with a puzzle – how can some fact about the psychology of a person be settled by appeal to evidence which does not concern the psychological states of the person, but rather evidence which concerns an independent world? Byrne divided the Puzzle of Transparency into three versions reproduced above: The Puzzle of Reliability, The Puzzle of Inadequate Evidence and The Puzzle of Reasoning through a False Step. We will examine how Byrne engages with each of these in turn.

5.4. The Puzzle of Transparency

5.4.1. The Puzzle of Reliability

Recall that the puzzle regarding the transparent inference was the following:

> How can reasoning in the pattern of the Doxastic Schema lead to reliably true beliefs? The premiss *p* could easily obtain without my believing it, so an argument in the form of the Doxastic Schema cannot prima facie be a reliable way to reach the truth.

Byrne expresses the puzzle of reliability in terms of epistemic rules like BEL as follows:

> "Yet surely this [BEL] is a bad rule; in other words, following BEL tends to produce false and unjustified beliefs. Putting it in terms of the first (reliability) variant of the puzzle of transparency: that p is the case does not even make it likely that one believes that it is the case." (Byrne, 2018, p. 103)

The question for Byrne at this stage is one of how following BEL can secure *knowledge* of what one believes. He suggests that BEL secures knowledge in virtue of certain features of the rule; BEL is *self-verifying*: "One is only in a position to follow BEL by believing that one believes that *p* when one has recognized that *p*. And recognizing that *p* is (inter alia) coming to believe that *p*." (Byrne, 2018, p. 104)

The combination of the schema by which one follows BEL, and the C-conditions, is sufficient to guarantee that if one follows BEL then the inference one makes is a good one. Should *S* follow BEL, *S* cannot fail to know that she believes that *p*, if steps (i)-(iv) of the schema are met. The schema not only guarantees that when *S* follows BEL she has a higher order belief, it also guarantees that such a transition is a reliable method for *S* to form their higher order beliefs, and that those higher order beliefs are true. That this transition is a reliable one is sufficient for *S*'s higher order belief to constitute knowledge of her first order belief. BEL's self-verifying nature ensures that whenever *S* follows BEL, the transition furnishes *S* with knowledge of her own beliefs.

5.4.2. The Puzzle of Inadequate Evidence

The Puzzle of Inadequate Evidence developed in terms of inference was the following:

> *P* is insufficient evidence for the conclusion that I believe that *p* – that it is raining is inadequate evidence for my belief that it is raining, but according to the schema, it is my total relevant evidence. So how can knowledge be based on this inadequate evidence?

The central thought here is that even if BEL is a reliable means of forming higher order beliefs, the premiss upon the basis of which one forms the higher order belief (the premiss *p*) is weak evidence that one has the first order belief regarding that evidence – the fact that *p* is poor evidence for the claim that I believe that *p*. Byrne appeals to a safety condition on knowledge to explain this apparent

puzzle – the contention is that BEL produces epistemically safe beliefs. Further, the status of one's self-knowledge as knowledge, gained by following BEL should not be explained in terms of an inference from evidence. Rather, the self-knowledge in question is "…like basic perceptual knowledge." (Byrne, 2018, p. 106)

Byrne takes it that safety, understood as "…one's belief that p is safe just in case one's belief could not easily have been false…"(Byrne, 2018, p. 110) is both a necessary and sufficient condition for knowledge, so one knows that *p* just in case one's belief that *p* could not have easily been false. Byrne asks us to consider again DOORBELL:

> If the doorbell rings, believe that there is someone at the door.

BEL is more epistemically secure than DOORBELL - recall that BEL is *self-verifying* so one's second order belief is guaranteed to be true. If, upon the ringing of the doorbell, one follows BEL, the following would obtain:

> (i) *S* Believes that she believes that the doorbell is ringing because she recognises that the doorbell is ringing.
>
> which implies
>
> (ii) *S* recognizes (hence knows) that the doorbell is ringing.
> (iii) The doorbell is ringing.
> (iv) *S* believes that she believes the doorbell is ringing.

S's belief that she believes the doorbell is ringing is guaranteed to be true simply in virtue of *S*'s following BEL. But what we are concerned with is not merely the truth of *S*'s higher order belief, but whether (as Byrne puts it), *S* could not have easily formed a false belief, i.e., is *S*'s higher order belief *safe*? Byrne considers three types of case in which one's belief about whether there is someone at the door could have gone wrong:

> "Type I: not-p, and one falsely believes that conditions C obtain, thereby believing that p. […]
>
> Type II: not-p, and one truly believes that conditions C obtain, thereby believing that p. […]
>
> Type III: not-p, and one believes that p, but not because one knows or believes that conditions C obtain." (Byrne, 2018, p. 110)

Of these, the follower of BEL is only vulnerable to type III errors – type I errors are ruled out because, recalling the formula for an epistemic rule – If conditions C obtain, believe that *p*. In Type I errors, *S* believes that *p* on the basis of falsely taking it that conditions C obtain, despite *p* not obtaining. BEL states that 'if *it is raining*, believe that you believe that *it is raining*.'

Substituting BEL into a type-I error:

> One does not believe *it is raining* and one falsely believe *it is raining* thereby believing that one believes *it is raining*

This is straightforwardly contradictory, and type I errors are ruled out.

Substituting BEL into a type-II error:

> One does not believe that *it is raining* and one truly believes that *it is raining* thereby believing that one believes that *it is raining*.

Again, the follower has contradictory beliefs, and this sort of error is ruled out.

The type III error, however, does not lead to an obvious contradiction:

> One does not believe *it is raining* and one believes that one believes that *it is raining*, but not because one knows or believes that *it is raining*.

This error is *odd* but the beliefs are not contradictory. The epistemic agent would have a second order belief which had no causal-rational connection to their first order belief, which is strange, but not straightforwardly contradictory[94].

Someone might ask, 'how does this lead to BEL producing beliefs which in the overwhelming majority of cases are true?' If the only errors one can make in even trying to follow BEL are type III errors, then one will very rarely err – type III errors are a remote possibility in the vast majority of cases (this is underlined by the seeming oddness of the type III error). The remote possibility of type III error, combined with the self-verifying nature of BEL leads to second order beliefs formed by subjects who follow BEL being safe beliefs. If a type-III error is the only error one who follows BEL is liable to fall into, then the only error one who follows BEL can fall into is irrelevant for determining whether second order beliefs are knowledge.

5.4.3. The Puzzle of Reasoning Through a False Step

Ordinarily, when following BEL, one reasons from a fact about the world, the fact that *p*, to the belief that one believes *p*. But this presumes that one does not make a mistake when directing one's eye out onto the world. Following BEL involves recognizing (hence knowing) that *p*. But one does not always get it right when one directs one's attention to the world. Byrne suggests that the way such failures on the part of the subject should be understood is in terms of *failing to follow* BEL – one

[94] One possible worry that type-III errors present is the case of a subject who pathologically made type-III errors. They would seem to be alienated from their mental life in a way not dissimilar to the self-blind individual described by e.g. Shoemaker (1996).

merely *tries* to follow BEL, but one does not succeed: "…S *tries* to follow rule R iff S believes that p because S *believes* that conditions C obtain." (Byrne, 2018, p. 107) One can try to follow an epistemic rule like BEL without successfully following it. One could, for example, see water coming past one's window (one's upstairs neighbours are watering their window boxes perhaps) and falsely conclude it is raining, then try to follow BEL and reach the belief that one believes it is raining. Our question is 'does this second order belief amount to knowledge that one believes it is raining?' Put in this way, the 'false step' version of the puzzle seems less puzzling. Why should it be puzzling that the output of even a failed attempt to follow BEL is a true second order belief that amounts to knowledge regarding one's first order belief (which itself may be false)? The status of the second order belief is not tied to the truth or falsity of the belief that *p*. If one believes conditions C obtain, one has the first order belief that *p* (which may or may not be a true belief). The second order belief is still guaranteed to be true simply in virtue of trying to follow BEL. As Byrne suggests, BEL is not merely self-verifying, it is *strongly self-verifying*. BEL is self-verifying even of one only tries to follow BEL but does not succeed. Like the objection from inadequate evidence, Byrne suggests that even in cases where one reasons through a false step, if one tries to follow BEL, one's belief is *safe*. Beliefs formed through BEL could not have easily been false, even if the first order belief one forms the higher order belief about is false. Trying to follow BEL, on very many occasions, will lead to self-knowledge.

5.5. Privileged and Peculiar Access Explained

5.5.1. Privileged Access Explained

Byrne's 'transition as inference' account of self-knowledge offers an explanation of privileged access. Recall that privileged access is the thesis that

> "…beliefs about one's mental states acquired through the usual route are more likely to amount to knowledge than beliefs about others' mental states (and, more generally, corresponding beliefs about one's environment)." (Byrne, 2018, p. 5)

The 'usual route' in Byrne's account is the following of BEL, and as suggested above, following (or trying to follow) BEL is more likely to amount to knowledge about one's own mental states, as BEL is strongly self-verifying and produces safe beliefs, than the corresponding method regarding the beliefs of others:

> "When you conclude that you believe that p from the premise that p, there is a causal transition between two states you are in: believing that p, and believing that you believe that p. The second belief state is true if you are in the first state. And since the transparent inference guarantees you are in the first state, this method is a highly reliable way of forming true beliefs." (Byrne, 2018, p. 109)

The method by which one knows the beliefs of others is not self-verifying, let alone strongly so, and so is less likely to amount to knowledge.

5.5.2. Peculiar Access Explained

Recall that peculiar access is the thesis that "…one can come to know about one's mental life "in a way that is available to no one else." (Byrne, 2018, p. 8). Again, BEL is strongly self-verifying, and S's belief that she believes *it is raining* is responsive to her belief that *it is raining*, not, for example, her interlocutor's belief that *it is raining*. As Byrne suggests "Inference is a causal process involving a single subject's mental states, which is why the transparency procedure is quite ill-suited to detect others' mental states." (Byrne, 2018, p. 109) Peculiar access is explained, the transparent procedure is not suited to producing beliefs about other's mental states.

5.6. A Lacuna: The Role of Reasons

In developing the account of what it is to follow an epistemic rule, recall that Byrne's suggestion was that the 'because' in the rule schema marked a 'reason-giving causal connection'. Given all we have said, and the responses to the three versions of the Puzzle of Transparency, it is not clear that the talk of reasons does significant philosophical work in Byrne's account. Rather, it seems to present a lacuna in Byrne's exposition. The invocation of reasons seems to play no significant part in Byrne's solution to the Puzzle of Transparency that could not be played by a causal connection, and likewise appears to play no explanatory role in the account of self-knowledge Byrne presents that could not be adequately covered by an invocation of only a causal connection. The first incarnation of the Puzzle of Transparency – the puzzle of reliability – is answered by appeal to a reliable causal mechanism, and the second and third incarnations are answered by appeal to the idea that BEL is self-verifying, a feature of the structure of the rule, not the role of the 'because' as reason giving. BEL would be equally as self-verifying if the 'because' marked only a causal connection between the recognition of C-conditions and the formation of one's higher order belief. Nothing in these answers requires talk of 'reason-giving' connections, and Byrne does not explicate the notion beyond a suggestion that a reason-giving causal connection would be akin to a 'basing relation', a further unspecified notion. Note that I do not (yet) present an objection to Byrne's position. Rather I suggest that more explanatory work must be done in order for us to make sense of the place of the 'reason-giving' arm of a 'reason-giving causal connection' in Byrne's dialectic. Indeed, as presented, Byrne's account can be understood as a causal-reliabilist account of self-knowledge which can proceed without needing to talk about reasons at all.

We can see this by re-examining BEL, but rather than the 'because' in the schematic epistemic rule marking a reason-giving causal connection, we can assume that it instead marks nothing more than a reliable causal connection (a regularity) between the antecedent and the consequent of the rule. Call this an *austere* reading of Byrne:

S follows the rule BEL ('If p, believe that you believe that p') on a particular occasion iff on that occasion:

(i) S Believes that she believes that p because she recognises that p

which implies

(ii) S recognizes (hence knows) that p

(iii) p obtains

(iv) S believes that she believes that p

So, on this causal reading of 'because', *S* is caused to believe that she believes it is raining by recognizing it is raining. The cause of *S*'s higher order belief is the recognition that it is raining. This rule is still self-verifying and is as such a reliable producer of true higher order beliefs. Further, each of Byrne's responses to the decomposed Puzzles of Transparency turn on BEL being self-verifying or producing safe beliefs. The causal transition BEL describes on this view is likewise a safe transition – it was argued above that the transition is safe because it is only vulnerable to type-III errors. This is equally true in the case where the 'because' in the epistemic rule is a causal transition – nothing in the rejection of type-I and type-II errors requires reason-talk. Type-I and type-II errors are ruled out because the beliefs that are formed in those sorts of errors are inconsistent – they involve a belief that p and also a belief that not-p, held together. This inconsistency is precisely the sort of thing that motivates adoption of the doxastic schema – the schema is supposed to rule out such inconsistencies as a 'Moorean absurdity'. That the inconsistency which motivates the doxastic schema can be ruled out by a causal-reliabilist account suggests that we also need not view Byrne's transition as an inference. That is, the transition described by the epistemic rule need not be a rational transition, a reliable causal transition will do. If Byrne wishes to talk of reasons, he must show how they are important to his account.

5.7. Byrne's Account and the Standard Objections to Transparency

Byrne does not engage directly with what I have called the Standard Objections to Transparency[95], save for his extended engagement with the Puzzle of Transparency. Nevertheless, we can still assess how Byrne's proposal engages with the objections.

[95] The Objection from Scope, The Objection from Over-Intellectualisation, The Anti-Luminosity Argument, and the Puzzle of Transparency. See section 1.4.

5.7.1. The Objection from Scope

The Objection from Scope suggests that accounts of self-knowledge based on the Transparency Remark have insufficient scope to be satisfactory accounts of self-knowledge. They either miss something out or are intentionally restricted accounts. Byrne's aim in *Transparency and Self-Knowledge* is to present a *unified* account of self-knowledge, where a unified account is in Byrne's terms an account that explains a subject's knowledge of her mental states or provides a general formula that can be extended to the subject's knowledge of any of her mental states. Although this chapter has focussed only on Byrne's account of self-knowledge of belief, this is because it is the central case for an account of self-knowledge; the account stands or falls based on what Byrne says about self-knowledge of belief. Byrne uses the account of belief as a foundation for accounts of perception and sensation, desire, intention and emotion and memory, imagination and thought. Further, the general formula of epistemic rules Byrne sets up in his development of the account of belief can be extended to other mental states if an explanation is needed. If Byrne is successful in giving his account of belief, the Objection from Scope is no objection to his account.

5.7.2. The Objection from Over-Intellectualisation

The Objection from Over-Intellectualisation suggests that an account which is developed from the Transparency Remark asks too much of a subject. As remarked in section 1.4.3., the objection puts little pressure on the Inferentialist, unless the inference which explains self-knowledge is understood as a conscious inference. It seems no inferentialist should accept the charge that the inference which explains self-knowledge is done consciously, and this is indeed Byrne's position.

Byrne's response to this is not merely that the inferences need not be conscious, they also cannot be 'critical reasoning' and that following the epistemic rule BEL is not self-intimating, i.e., it can be done without one knowing one is doing so. If Byrne's Transparent transition is unconscious, the Objection from Over-Intellectualisation has no grip at all on his account of self-knowledge.

5.7.3. The Objection from Anti-Luminosity

Byrne does not engage directly with the Anti-Luminosity Argument, and this is, I suggest, because Byrne does not accept the luminosity claim (L):

> "(L) For every case *a*, if in *a*, C, then in *a* one is in a position to know that C obtains." (Williamson, 2000, p. 95)

Byrne's position does not claim that by following an epistemic rule, (L) can be vindicated, indeed it seems that his position is a denial of (L). Focussing on the belief case, Byrne's claim is not that belief is luminous (which we might understand as the following: for every case where a subject believes *p*, she knows (or at least is in a position to know) she believes *p*). Byrne's claim is rather that by

following BEL, a subject's belief that she believes *p* is true, because following BEL is a self-verifying rule. This is not to say that whenever a subject believes *p* she has a belief about her belief, but rather that, if a subject does have a belief about her belief (formed by following BEL), that belief is true. This is not the claim in (L). Byrne need do nothing to fend off the Anti-Luminosity argument.

5.8. A Further Objection to Byrne's Account – Byrne against Evans

Byrne's position is that the 'transparency' of the question 'do you believe that *p*?' to the question 'is it the case that *p*?', drawn from Evans' Transparency Remark, should be understood as an inference, and that making this inference consists in following a special sort of rule, an *epistemic rule*. The chapter thus far has provided an exegesis of Byrne's view. In this section, I intend to show that given a plausible constraint on inference clearly endorsed by Evans, Byrne's account fails to be a development of Evans' remark. Indeed, if the objection I present has teeth, Byrne fails to appreciate the significance of Evans' remark in such a way that while he might still present his account as an Inferentialist account of self-knowledge, he does nothing to suggest that it might be preferable to an account which does develop Evans' remark, such as the account offered in this thesis.

The objection I present proceeds quite simply: I will suggest that if Byrne accepts a plausible condition on a transition being an inference, the Why-Condition, then he must deny Evans' insight from the Transparency Remark. If Byrne denies this, his account cannot in any sense be an explication or development of Evans' thought. Byrne might still suggest his account is the correct Inferentialist account of self-knowledge even still, but given he takes his position to be an explication and extension of Evans' view, he has done nothing to argue against Evans' view such that we might prefer his. Evans' view is simply not in focus for Byrne and he must say more to discount it.

A plausible constraint on inference is that a subject performing an inference is in a position to answer a certain sort of question, a 'why' question relating to their conclusion: 'Why is the conclusion the case?' Call this constraint the Why-Condition.

If a subject infers the conclusion *q* from the premises '*p*' and '*p* implies *q*', they are, it seems, in a position to answer the following question: 'why *q*?'. The subject can answer this question by asserting 'because *p* and because *p* implies *q*'. Take a concrete example: Alice looks out of her window and sees that it is raining. From her seeing that it is raining and the premiss that if it is raining, the ground is wet, she infers that the ground is wet. She opines to her officemate Bob that the ground is wet, and Bob asks her why the ground is wet. By performing the inference from 'it is raining' and 'if it is raining, the ground is wet' to 'the ground is wet', Alice is in a position to answer

Bob; she can point to the premises of her inference as answers to a certain sort of 'why' question[96]. She can answer the question 'why is the ground wet?' with the answer 'the ground is wet because it is raining, and if it is raining, the ground is wet'. A subject who has performed an inference is in a position to answer the question of 'why the conclusion?' The Why-Condition makes explicit something we take for granted in understanding inference: that in making an inference the subject understands the logical relationship between a pair of contents – the premiss and conclusion of the inference. This is not to say that the subject has explicit logical understanding of a rule of inference (e.g., Modus Ponens), but in making an inference, the subject is in a position to understand that the first content follows from the second content. It is internal to inference that if the subject is asked why the content which forms the conclusion obtains, she can answer by reference to the content that forms the premises. Were she not in a position to answer the why-question, it is unclear by what lights we can say she inferred her conclusion. We might equally say the conclusion popped into her head or she concluded it based on nothing at all. Thus, the Why-condition is not an external constraint imposed upon a transition to make it a good inference or the like but is rather an explication of the internal connection between the content which forms the premises and the content which forms the conclusion of an inference, a connection the inferring subject is sensitive to in making her inference.

I have not yet said, however, how the Why-Condition relates to Evans' remark. The connection is made clear in the important place given to answering questions in the Transparency Remark:

> "If someone asks me `Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?'" (Evans, 1982, p. 225)

Evans' formulation of the Transparency Remark is in terms of questions and their answers. From this formulation, we can take the following thought: to believe 'there will be a third world war' is to be able to answer the question 'will there be a third world war?'; to believe *p* is to be able to answer the question 'is it the case that *p*?'. This is implicit in Evans' formulation of the Transparency Remark. It can appear, however, that being in a position to answer the question 'is it the case that *p*?' is not the same as being in a position to answer the question 'why is it the case that *p*?' (or 'why *p*?' for short), and the question 'why *p*?' is the question relevant to the Why-Condition. But reflecting on Evans' remark shows us the relationship between the questions. Answering the question 'is it the case that *p*?' in the canonical fashion involves answering the question 'why is it the case that *p*?' This

[96] It seems natural to talk of Alice's *reasons* here. We should, I think, resist this, unless all we mean by 'reasons' is 'answers to a certain sort of question'; Evans talks of 'being in a position to answer questions' but not reasons and moving to talk of 'reasons' brings in significant philosophical baggage which would only obscure the simple point being made here.

is, as suggested in chapter 4, the insight of the deliberative account. Answering the question 'is it the case that *p*?' in the canonical fashion puts one on a position to answer the question 'why is it the case that *p*?'

To see how the endorsement of the Why-condition is incompatible with Byrne's account being an extension of Evans', we must recall that what Evans' Transparency Remark tells us is that (as Evans puts it) "…whenever you are in a position to assert that *p,* you are *ipso facto* in a position to assert 'I believe that *p*'." (Evans, 1982, p. 225-6). Just so, whenever you are in a position to answer the question 'why *p*?' you are in a position to answer the question 'why do you believe that *p*?': Evans' point is that to answer the question 'do you believe *p*?' I need do nothing more than I would in answering the question 'is it the case that *p*?' – the *very same procedure is put in place to answer to both questions*. As such, it seems that to answer the question 'why do you believe that *p*?' one would put into operation the very same procedure as in answering the question 'is it the case that *p*?'.

Given this link between the why-questions, let us return to the case of inference. Evans' remark suggests that a subject who infers *q* from the premise's '*p*' and '*p* implies *q*' is in a position to answer both the question 'why *q*?' and the question 'why do you believe *q*?'. Surely one who has inferred *q* from *p* and '*p* implies *q*' can answer the question 'is it the case that *q*?' – they have inferred *q*, so they *must* be in a position to answer this question. But this means that by Evans' lights, anyone who has made this inference, and as such is in a position to answer the question 'is it the case that *q*?' can answer the question 'why do you believe *q*?'. That is, if the subject *S* has inferred *q*, she can affirm that she believes *p* and that she believes that *p* implies *q*, and so she believes *q*. The inferring subject who concludes *q* (i.e., is in a position to assert 'it is the case that *q*') can explain her belief that *q* in terms of her belief that *p* and her belief that *p* implies *q*; she can answer the question 'why do you believe *q*?' with the answer 'because I believe *p* and I believe *p* implies *q*'. So, the subject who believes *q* (is in a position to assert 'it is the case that *q*') and can answer the question 'why *q*?' thus knows that she believes *p* and knows that she believes that *p* implies *q*. Evans' remark suggests that there is an internal connection between the belief in the conclusion of an inference and the belief in the premises.

However, Byrne can accept the Why-Condition only if answering the question 'why *q*?' does *not* put the subject in a position to answer the question 'why do you believe *q*?', for if it did, the inference would be self-intimating. Byrne must deny that the capacity for inference implicates or requires self-knowledge on the part of the inferring subject, or his account is no explanation of self-knowledge. Byrne aims to explain a subject's self-knowledge in terms of her capacity to make inferences

(inferences which are described by epistemic rules), and if the capacity to make inferences requires self-knowledge on the part of the subject (e.g., the knowledge that she believes *q* because she believes *p*, or that she knows she believes *p* as part of the inferential process of concluding *q*), then the capacity for inference cannot be the explanation of self-knowledge as the capacity for inference assumes the very self-knowledge it aims to explain.

Byrne is, of course, sensitive to this. In characterising inference as the following of an epistemic rule, Byrne characterises the rule as *neutral*. A neutral rule cannot implicate self-knowledge in its antecedent, and further, he is explicit that the following of such a rule is itself not self-intimating:

> "There should be no temptation to think that rule-following is self-intimating: one may follow a rule without realizing that this is what one is doing." (Byrne, 2018, p. 102)[97]

As such, Byrne cannot accept the connection between the two 'why-questions' which falls out of Evans' insight that answering the question 'is it the case that *p*?' puts one in a position to answer the question 'do you believe that *p*?'. To do so would be to say that inference is self-intimating and his account is no account of self-knowledge. Byrne is presented with a trilemma:

On the first horn, Byrne would accept that inference is self-intimating, and he gives no account of self-knowledge. This is obviously intolerable.

On the second horn, Byrne would need to deny the very position he takes himself to be explicating. To hold on to the Why-Condition, he must concede that his position cannot not be a development of Evans', or a fuller understanding of Evans' remarks, but must instead be their denial. Byrne's 'Transparency' account would, in a sense, be no such thing. That is not to say that it would not give an account of self-knowledge, but that account would not, in the relevant sense, be descended from Evans'. Of course, Byrne might accept that his account is not in the relevant sense a development of Evans' remark, but is still the correct, Inferentialist, account of self-knowledge, and as such, he can hold on to the Why-Condition. If so, then Byrne has done nothing to suggest why we might prefer his Inferentialist account of self-knowledge over an account which does develop on Evans' thought, an account I hope I have made plausible.

On the third horn, Byrne could deny the Why-Condition outright and deny that his account is an inferentialist account of self-knowledge, but instead is a causal-reliabilist account of self-knowledge of the sort suggested in section 5.6. To do this would also free him from answering the lacuna regarding reasons, but would, I suggest, be ultimately as intolerable as accepting that inference is self-intimating.

[97] Note that Byrne suggests that there are self-intimating epistemic rules, but the key point is that the belief rule BEL, the central rule for self-knowledge *cannot* be self-intimating on pain of circular explanation. Rules for the explanation of other mental states can be self-intimating as long as the central case is not.

5.9. Concluding Remarks to Chapter 5

In this chapter I have provided an in-depth discussion of the central case of Byrne's explication of the Transparency Remark. Byrne's discussion represents the best option for a unified Inferentialist account[98] which aims to be a development of Evans' remark. I have suggested that Byrne's account cannot be both inferential and a development of Evans' remark, and that there is a significant lacuna in the account regarding the role of reasons in Byrne's dialectic. This, I suggest, at best leaves it open whether the account is successful by Byrne's own lights. I suggest that there are two live options for Byrne. He could, suggest that his account is no development of Evans' account, but then as suggested above, Byrne would need to do more to motivate why we should accept his Inferentialist story given the robust development of Evans' remark in this thesis. As it stands, Byrne has done nothing to suggest that an account which properly develops Evans' point should be discounted; he has provided no objections because he took his account to be a development of Evans. Further, if Byrne accedes and takes the route of denying his account is a development of Evans, but nevertheless it is still inferential, the lacuna regarding reasons discussed above still remains. Of course, this is not a knock-down refutation of Byrne's position, but he is left with philosophical work to do. Alternatively, perhaps Byrne could retreat to a causal-reliabilist account which drops talk of inference as anything other than an epistemically significant transition. This retreat would remove the lacuna regarding reasons, but such an account would *still* be no account of Evans' remark, it would rather be a well-developed causal-reliabilist account of self-knowledge based around the idea of a strongly self-verifying transition. To retreat in this way would be a significant concession in aim for Byrne's account, but would garner no further philosophical cost than a reduced ambition. Regardless of which of these options Byrne chooses, given that Byrne's account is the best developed unified Inferentialist account, I suggest that I have at the very least made the Rationalist alternative offered in this thesis palatable as a development of the Transparency Remark and a first step in an account of self-knowledge in general.

---

[98] Cassam (2014) presents an inferentialist account of self-knowledge which denies unification. Cassam does not aim to explain all self-knowledge via Transparency or inference:

> "Another worry about inferentialism might be that it goes against my insistence that when it comes to explaining human self-knowledge there is no magic bullet, no one source that is capable of accounting for all our intentional self- knowledge. Isn't inference a single source? There are two things to say about this: *first, saying that inference is a key source of intentional self-knowledge for humans doesn't mean that there aren't other sources*. […] The second point is that 'inference' as I understand it is such a broad category and covers so many different things that inferentialism hardly amounts to a 'magic bullet' explanation of human self-)." (p. 140) [emphasis mine]

As such, Cassam's inferentialist account is no alternative to the Simple Account; it does not meet the Objection from Scope.

6. Final Remarks

6.1. Some Remarks on The Simple Account and Alternative Accounts of Self-knowledge

In Chapter 1, I provided a review of some other Rationalist developments of the Transparency Remark, and the Simple Account shares some features with those accounts. In those sections, I did not detail how the Simple Account differs from those positions. Here I will suggest how the Simple Account is related to *constitutivist* accounts of self-knowledge, and how the account differs from Finkelstein's expressivist account of Transparency.

6.1.1. The Simple Account and Constitutivism

In section 1.3.3. I suggested that Marcus and Schwenkler's (2018) 'self-consciousness conception' of the Transparency Remark was ultimately a form of constitutivism. We might ask now whether the Simple Account, which shares some of the features and motivations of Marcus and Schwenkler's position is itself ultimately a constitutivist position. The answer to this question turns on how tightly we delineate the constitutivist position. In that section, recall that the suggestion was that the constitutivist is committed to a biconditional of the following form:

> "Constitutive Thesis: Given C, one believes/desires/intends that P/to Φ iff one believes (or judges) that one believes/desires/intends that P/to Φ." (Coliva, 2016, p. 164)

Marcus and Schwenkler, however, suggested a considerably weaker form of constitutive thesis:

> "Constitutivist theories hold that we can have non-empirical knowledge of our beliefs because to take oneself to believe something is, at least in the right conditions, also to believe it." (Marcus and Schwenkler, 2018, p. 15)

We might understand Marcus and Schwenkler's suggestion as pointing to a liberal form of constitutivism, where the claim is (ultimately) that there is some important connection between the belief 'p' and the belief that (as each of us would put it) 'I believe p' . The liberal constitutivist is, on this view, any theorist who is committed to some important connection between the beliefs in question but does not endorse the biconditional. The job of the theorist on the liberal picture is to explicate the 'important connection' between the beliefs in question such that they give an account of self-knowledge of belief.

Coliva's reconstruction of the constitutive thesis, on the other hand, is considerably stronger. Call the theorist who endorses the biconditional account Coliva suggests a strict constitutivist. The business of the strict constitutivist is, as suggested in section 1.3.3., to explain the conditions under which the relation of constitution holds (the C-conditions), and to explain whether and why one direction of explanation of the biconditional takes priority over the other (or that there is no

priority). Note that the explanatory demands placed on the strict constitutivist are, at least initially, more demanding than the liberal. The strict constitutivist already has in place an account of the relation between beliefs – it is a constitutive relation, and they take it what remains to be explained is how, and under what conditions, that constitutive relation can be applied to the states in question.

The liberal constitutivist, by contrast, has what seems to be a looser explanatory demand – they must explain the 'important connection' between the beliefs in question in such a way that whatever the connection is, it delivers (or underwrites) an account of self-knowledge of belief.

With this ground-clearing in place, let us look again at the Simple Account. The Simple Account does endorse an 'important connection' between the belief 'p' and the belief 'I believe p', and as such the Simple Account is, by the suggested taxonomy at least a liberal constitutivist account. Further, the Simple Account does tell us what that important connection amounts to – the idea that the very same epistemic capacity is exercised in having the belief 'p' as in having the belief 'I believe p'. The connection suggested is also sufficient to deliver an account of the self-knowledge of belief. Recall that the strict constitutivist suggests that as long as certain conditions are met (as long as C-conditions obtain), there is a biconditional relation between the belief 'p' and the belief 'I believe p'. The Simple Account also endorses such a position – asserting 'p' puts one into a position to assert 'I believe p', iff the assertion 'I believe that p' is (would be) made on its canonical basis. That is, the C-condition that regulates the obtaining of the biconditional between the belief 'p' and the belief 'I believe p' is that the belief 'I believe p' is formed on its canonical basis. Condition C obtains iff the belief 'I believe p' is formed on its canonical basis, and as such the biconditional obtains iff the belief 'I believe p' is formed on its canonical basis. So, the Simple Account satisfies one explanatory demand for a constitutive account, but I have not yet answered the explanatory demand regarding direction of explanation, nor have I shed any light on what the relation of constitution might amount to. To answer the demand regarding direction of explanation, we should (although I suggested above that the strict constitutivist perhaps need not) give an account of what is constituted by what in the Simple Account. Evans' remark suggests that the question 'do you believe p?' can be answered by doing nothing more than answering the question 'is it the case that p?'. As such, it seems appropriate to say that answering the question 'is it the case that p?' is constitutive of answering the question 'do you believe that p?'. This is not to say that answering the question 'is it the case that p?' simply is answering the question 'do you believe that p?', that is what the locution of 'putting one into a position to answer' clarifies. Answering the question 'is it the case that p?' constitutes an answer to the question 'do you believe that p?' because the subject need do nothing more epistemically to answer the question of belief than answer the world directed question. The

questions are answered by exercise of the very same epistemic. So, it is not so simple as the belief 'p' constituting the belief 'I believe that p' given condition C is met, as a flatfooted reading of the constitutive thesis might suggest. Rather, Evans' locution of answering questions provides illumination of what the constitutive thesis might amount to; the answer to one question constitutes an answer to the other because the canonical basis of the answer to the question 'do you believe that p' is the very same basis as that of the answer to the question 'is it the case that p', i.e. the questions are answered by exercise of the same epistemic capacity. This suggests that iff the question 'do you believe that p?' is answered on its canonical basis, a biconditional holds between the answer to the question of belief and the answer to question of what is the case.

We are now in a position to suggest the explanatory priority of the biconditional given the understanding of the constitutive relation suggested. The canonical basis of the belief 'I believe p' is the basis of the belief 'p', and the answer to the question 'is it the case that p?' is in some sense 'prior' to the answer to the question 'do you believe that p', insofar as the answer to the latter is constituted by an answer to the former. So, the direction of explanation is from the belief 'p' to the belief 'I believe p', i.e., a left-right direction. A rejoinder might be that I can surely, by this biconditional, find out the state of the world by reflecting on what I believe, and that's mad. And the answer is yes, but it isn't mad – one can come to know what one believes (an answer to the question 'do you believe p?' via the Simple Account only if one comes to know what one believes via its canonical basis, i.e. whatever basis one believes p on, so by investigating what I believe (as long as I do so in the right way) is inter alia investigating the world (and vice versa). That is, it isn't mad as long as condition C is met.

With all this in place we might now challenge the import of the Simple Account. If the Simple Account is nothing more than a strict constitutivist account, why prefer it over other constitutive accounts? I have not directly engaged with other constitutive accounts here, but the Simple Account obeying the constitutive biconditional should be no mark against it. Rather, the Simple Account shows that the Transparency theorist and the Constitutivist may not be so far apart. That the constitutive thesis is a consequence of following through Evans' reasoning to its completion is under-appreciated, and the Simple Account shows that it is at least a compelling consequence of a rationalist reading of Evans. Further, the Simple Account satisfies the explanatory demands of the constitutivist and gives a principled explanation of how we might understand the relation of constitution, and as such is an independently interesting account of at least self-knowledge of belief, regardless of how it performs compared to other 'pure' constitutive accounts.

6.1.2. The Simple Account and Finkelstein's Expressivism

Recall that in section 1.3.3. I suggested that Finkelstein's account of Transparency (Finkelstein, 2012) gets something right. We are now in a position to work through what that is and see how the Simple Account is distinct from Finkelstein's expressivist account. Recall that Finkelstein's position is that subjects learn to use the locutions 'I believe that p' and 'p' interchangeably; if the assertion 'p' is an expression of the speaker's state of mind, then so is the assertion 'I believe that p': unlike the assertion 'p' the latter is a report of the speaker's state of mind; but Finkelstein's point is that it is not merely a report of this state, but equally an expression of it. The assertion 'p' is (for Finkelstein) an expression of the judgement or belief that p. But the speaker can also express her belief by asserting 'I believe that p' – her self-ascription of her belief is interchangeable with her expression of the same belief. Finkelstein takes Evans' Transparency Remark to be a beginning move toward such a position – the reason one is in a position to answer the question 'do you believe that p?' on the basis of answering the question 'is it the case that p?' is because the speaker learns to use the answers to these questions, the expressions 'I believe that p' and 'p', interchangeably. Finkelstein then explains Authority by suggesting that the expression 'I believe that p' expresses the very belief self-ascribed, so directly expresses the speaker's psychological state, in the same way that the self-ascription 'I am in pain' expresses a speaker's pain as directly as a groan. The self-ascription is an expression of the very attitude self-ascribed. Only the speaker is in a position to express the attitude she self-ascribes, so we have an explanation of Authority.

In many ways what Finkelstein suggests is congenial to the Simple Account, but there are important points of departure. Finkelstein is on to the thought that the assertions 'p' and 'I believe that p' are made by exercise of the same capacity – that is one way to understand the idea that they can be used interchangeably by the speaker, but it is not the only way. There is nothing incoherent about a position that suggests that one can express their belief 'p' by the exercise of one capacity and self-ascribe that belief by exercise of another. All Finkelstein needs is the idea that the two expressions are used interchangeably in very many contexts – he need say nothing about their aetiology. As long as the capacity employed in the self-ascription of belief does not amount to an 'inward glance' – a detection by the subject of what she believes then an ascription of what is detected, Finkelstein can allow that different capacities are exercised in the expression 'I believe that p' and the expression 'p'. The expression of the Authority of avowals of belief is explained in a different manner by Finkelstein than the explanation offered by the Simple Account. As suggested above, Finkelstein's explanation of Authority is that the self-ascription of belief expresses the very same state of mind that it self-ascribes, so the subject is authoritative in her self-ascriptions. The Simple Account, by contrast suggests that Authority is explained by the fact that the assertion 'p' and the assertion 'I

believe that p' are made by exercise of the same epistemic capacity. Recall that the suggestion in 3.4.2 and 3.4.3. was that if a subject asserts 'p' on the canonical basis of that assertion, it is guaranteed that her assertion 'I believe that p' will be true, and it is guaranteed irrespective of whether her assertion that p is true. Her assertion 'p' does not need to be true; it is not any more secure even on this canonical basis, but the Transparent nature of belief guarantees the truth of her higher order belief. Her assertion regarding her own belief is in better epistemic shape than one regarding another of because it is guaranteed to be true.

The significant point of departure between the Simple Account and Finkelstein's expressivist account is in the treatment of the Puzzle of Transparency discussed in sections 3.2 to 3.4. The substance of the puzzle was very roughly in making sense of how the assertions 'p' and 'I believe that p' could be made by exercise of the same epistemic capacity but have distinct truth conditions. Finkelstein is presented with an analogous puzzle – if the expressions 'p' and 'I believe that p' are interchangeable insofar as they express the same thing, how can they have distinct truth conditions? Finkelstein doesn't engage much with this challenge in From Transparency to Expressivism, although he does suggest some sensitivity to it:

> "Why, in spite of their having different truth-conditions, am I able to use (e.g.) "It's about to rain" and "I believe it's about to rain" interchangeably? The answer could be put this way: It's because (in most circumstances) I'll express the same attitude – the same belief- regardless of which of these sentences I utter." (Finkelstein, 2012, p. 114)

Finkelstein seems to draw a distinction in this passage between that which the utterances express and the truth-conditions of the utterances. I suggest that in saying this, Finkelstein is in effect not engaging with the challenge – for to engage with the challenge would be to ask how the utterances can be on par with respect to what they express while having distinct truth-conditions, precisely the sort of challenge The Puzzle of Transparency presents, and understands in terms of, the idea that the utterances express (what we have called) 'the same general content'. There are ways in which we could unpack Finkelstein's position such that he can provide an answer to the Puzzle, but I need not do so here – I need only draw attention to the difference in engagement with the Puzzle between Finkelstein's account and the Simple Account.

6.2. Achievements of this Thesis

The central achievement of this thesis is a novel account of a subject's self-knowledge of her belief which develops the Transparency Remark in a way that both remains true to Evans' original motivation while answering the central objections to Transparency. This novel account, the Simple Account, in answering the Puzzle of Transparency, reveals the deep connection to Anscombe's work on the first person. Further, in understanding how the Simple Account both relates to 'deliberative

accounts' of Transparency and how the Simple Account might answer the Objection from Scope we have revealed how the Simple Account tells us something fundamental about accounts of self-knowledge, and how the form the Simple Account suggests might be marshalled to provide a general account of self-knowledge. I have also suggested how the best developed Inferentialist alternative to the Simple Account, Byrne's account, can be challenged via an internal objection that the account does not ultimately achieve what Byrne intended it to. Nevertheless, there is still work to be done to develop the Simple Account into domains of self-knowledge beyond those of belief, and one remaining question the Simple Account must answer.

6.3. Remaining Questions and Future Work

Given all that I have said in this dissertation, there are still two significant remaining questions and areas for further expansion, which have only now come properly into focus. The first of these is the question of how far the Simple Account can extend, given the comments on the Objection from Scope in at the end of chapter four. The second is Evans' contention that accepting Anscombe's point leads to an 'idealist conception of the self'.

6.3.1. The Generality of Transparency

The thought canvassed at the end of chapter four was that the Objection from Scope is defeated by focussing not on the content of the Simple Account of Transparency, but on the form of the account. There I suggested the Simple Account gives a story about the self-knowledge of belief and that the form of this account gives a general form that an account of self-knowledge in a domain could take:

> To have self-knowledge of being *F*, the subject need do nothing more than exercise the capacity the exercise of which is her being *F*.

This is the fundamental form of the Simple Account. In the Simple Account, we replace 'being *F*' with 'believing that *p*', and we have a way of understanding Evans' Transparency Remark. Of course, simply giving a way to understand the remark was not enough, and this thesis has attempted to carefully explicate this formulation in such a way that the consequences of Evans' view are in focus. This careful spelling out of what the general insight applied to a domain looks like, provides a model for how to approach (at least) the central domains of self-knowledge. I suggested in that chapter that this project is beyond the scope of this thesis, because on this general account of Transparency, an account of self-knowledge in a domain is simply an account of that domain. An account of a subject's knowledge of her intentions (as suggested in chapter four) would be nothing more than an account of intention (or practical reason)[99]. Likewise for sensation, desire, hope, judgement, or any

[99] We can see an account of intention which is similar to this general form in Anscombe's (1963) *Intention*. In particular, the idea of 'knowledge without observation' and the formulation of intention in terms of the

other attitude or state we might think comes under the scope of self-knowledge. This is the work of a general research project to extend an account with the general form across domains and is as such beyond the scope of this thesis[100]. The generality on offer also prompts the question of what general features of self-knowledge can be clarified based on this account. In chapter three I suggested that the Simple Account suggests that knowledge of one's beliefs is no cognitive achievement, insofar as one does nothing more than believe to be in a position to have knowledge of one's beliefs. If this is a general feature of self-knowledge, what does it mean to say self-knowledge is ultimately no cognitive achievement? I suggest that this points to a radically deflationary account of self-knowledge. In a sense, self-knowledge of belief is nothing more than beliefs about the world; it 'comes for free' in virtue of having beliefs. This radical proposal needs more careful development to understand what the ultimate consequences are for our understanding of ourselves as knowing subjects. Further, the Simple Account provides an explanation of the authority and groundlessness of self-knowledge of belief, and the general formula suggests that this explanation can extend to other domains, but it is not implausible that more may need to be done to understand the upshots of this authority and groundlessness for self-knowledge in general, although as noted in section 4.4., the generalisation of the Simple Account to sensations could present a particular challenge which may have significant philosophical consequences. The extension of the Simple Account to self-knowledge in general is the first major area of expansion of the ideas within this thesis. The second is a puzzle left to us from Anscombe's view, which is ultimately, I think, inseparable from the general conclusions we might draw from the Transparency of self-knowledge.

6.3.2. The Puzzle of the 'Idealist Conception' of the Self

The final puzzle is one that Nagel, and Anscombe, leave us. As suggested above, I will not attempt to solve this puzzle, instead I will try to give focus to it, without succumbing to what Anscombe might call 'raving'[101].

applicability of questions. Of course, this is barely even an attempt at a sketch of how Anscombe's work might be used to understand the transparency of intention. The aim is to suggest that there may be a fruitful starting point to understanding such an account in *Intention*.

[100] Kern's (2017) *Sources of Knowledge* provides an account which might be brought under this general form to as an account of self-knowledge of perception. Like the note regarding *Intention*, however, this is a promissory note rather than an attempt to integrate Kern's discussion into a general account of self-knowledge.

[101] "With that thought: "The I was subject, not object, and hence invisible", we have an example of language itself being as it were possessed of an imagination, forcing its image upon us.
The dispute is self-perpetuating, endless, irresoluble, so long as we adhere to the initial assumption, made so far by all the parties to it: that "I" is a referring expression. So long as that is the assumption you will get the deep division between those whose considerations show that they have not perceived the difficulty - for them "I" is in principle no different from my "A"; and those who do - c would - perceive the difference and are led to rave in consequence." (Anscombe, 1981, p. 32)

Evans suggests the puzzle is this: If we accept Anscombe's argument that assertions with the first-person pronoun as subject do not implicate identifying knowledge on the part of the speaker, we are left with what he calls an 'idealist conception of the self', the idealist conception being "…the same as saying that 'I' does not refer to anything." (Evans, 1982, p. 212, fn. 14) Evans reaches this conclusion through two related concerns. First that we are fundamentally *persons*[102] understood as elements of an objective order:

> "[O]ur thoughts about ourselves are about *objects* – elements of reality. We are, and can make sense of ourselves as, elements of the objective order of things." (Evans, 1982, p. 256)

Second, that this idea that our thoughts about ourselves are about objects entails that "[o]ur thinking about ourselves conforms to the Generality Constraint." (Evans, 1982, p. 256).

The 'Idealist conception' Evans wishes to guard against is the idea that my thoughts about myself are not thoughts about objects conceived as thoughts about an element of the objective spatio-temporal order. It is easy to see, I think, why Evans might conclude that if 'I' is not a term of reference, thoughts with 'I' as subject, are not thoughts about an object. Reference, on Evans' view is the singling out of an object from the manifold in such a way that the one who singles the object out knows which object she is singling out, i.e., she has identifying knowledge of an object. If 'I' is not a term of reference, if assertions with 'I' in the subject position do not implicate identifying knowledge of the subject, Evans suggests there is no *object* that the thought is about[103]. If this is so, if there is no object that my thought of myself is about, then such thoughts cannot, Evans suggests, fall under the Generality Constraint. Assertions with 'I' as subject would not be general in the way assertions with an object referring singular term in the subject position are. But we discussed in chapter three how we might retain a notion of generality even in the face of a no reference view of 'I'. As such the puzzle cannot be one of how an assertion with 'I' in the subject position which does not implicate identifying knowledge can be appropriately general. Rather the puzzle must be the puzzle of there being no object that I think of when I think 'I'. But what does this thought even amount to?

Nagel approaches a formulation of the puzzle in *The View from Nowhere*, asking the following: "What kind of fact is it – if it is a fact – that I am Thomas Nagel? How *can* I be a particular person?" (Nagel, 1986, p. 54). We can see the link between Nagel's remark and Evans' concern by recalling that to be a particular person here is to be an object. If my use of 'I' doesn't refer to an object, what

[102] Recall that the identification ⌜ I= δI ⌝ is the fundamental identification of a person.
[103] Evans calls this Anscombe's "…extraordinary conclusion that self-conscious thought is not thought about an object at all—that the self is not an object." (Evans, 1982, p. 214)

do I think of when I think 'I'? How can it be that I am a particular person, a particular object in the world?

Nagel decomposes this question into two related questions:

> "[T]he first half of the question is this: how can it be true of a particular person, a particular individual, TN, who is just one of many persons in an objectively, centreless world, that he is me?
>
> The second half of the question is perhaps less familiar. It is this: how can I be *merely* a particular person? The problem here is not how it can be the case that I am this one rather than that one, but how I can be anything as specific as a particular person in the world at all – any particular person." (Nagel, 1986, p. 55)

We might think that Nagel's puzzlement arises from a sensitivity to the following question: what is the connection between the 'I' of the transcendental unity of apperception (the 'I' of the 'I think' which accompanies all our representations), and the 'I' that expresses the empirical concept of a subject of experience? There is something inapt about this formulation, however. Putting the question this way suggests that there are two uses of 'I', one which is 'formal' or non-referring, and another which refers to a particular person. But this is not Anscombe's point. There are not two uses of the first-person (two 'I's'), one of which is object identifying, the other of which is not. There is one use, and it does not implicate identifying knowledge on the part of the subject, but nevertheless still conforms to the truth-conditional reference rule ''I am F' is true just in case the speaker is F'. The puzzle is understanding the consequences of Anscombe's argument.

The challenge is not to reconcile two 'I's'. It is the question of what sort of statement is 'I am NN', if not an identity statement? What is the connection between the empirical object within the spatio-temporal order that speakers would identify as 'NN'[104], and the 'I' of self-consciousness? What settles it that my non-identifying use of 'I' and my identifying use of a name are talking about the same thing? The clearest expression of this puzzle is developed in Haddock (2019):

> "No identifying knowledge is internal to any use of "I". But then it seems that, from the standpoint of what is internal to a use of "I", what "I" expresses does not concern anyone at all. Of course, it is possible from outside of this standpoint to assign a truth condition to a use of a sentence with "I" as subject, by means of Anscombe's rule, and in this light to count the use of the sentence as concerning the one who figures as the referent of the device in subject position in the specification of this condition. But that is to proceed from the external perspective of a theorist; from the standpoint of the self-consciousness that is internal to the use of "I", what "I" expresses does not concern anyone at all. Or so it seems. But then it is a real question what sense can be made, from this standpoint, of what is expressed by a sentence such as "I am this body" or "I am NN"." (p. 967)

[104] It is tempting to say 'the empirical object that I am that speakers identify as 'PC', but of course this already presupposes a connection between the non-identifying 'I' of self-consciousness and the identifying name 'PC', and as such is an illicit formulation

These questions all attempt to articulate the same puzzlement Nagel demonstrated above, and they leave us in a position to identify why this puzzlement should concern us.

The real worry the Idealist Conception generates is that if we have nothing more than an Idealist Conception of the self, then from within self-consciousness, i.e., from within what is expressed by the use of 'I', there is no object present or given to the mind *at all*. But if this is so, then consider the sentence 'I am NN', an (apparent) identity sentence. From within what is expressed by 'I' in this identity statement, there is nothing there – no object is present to the mind. But from within what is expressed by 'NN' in the identity statement, there is someone there – a particular person, NN. It seems to be inseparable from the idea that NN is a particular person that what is present or given to the mind in the use of 'NN' is a particular person. If nothing is given to the mind in the use of 'I', then it seems we must reject that (as each of us would put it) 'I am a particular person', for no person is present or given to the mind from within the use of 'I', so we have Nagel's question of 'how I can be a particular person?', and the question of how can 'I am NN' be an identity statement, and if it is not, what sort of statement is it?

Anscombe's argument holds a fundamental place in properly understanding an account of self-knowledge developed from the Transparency Remark, and armed with this understanding, we must (I think) engage this further puzzle. That we must understand self-consciousness by engaging with this puzzle is, however, no objection to the Simple Account. Evans' worry that if we accept the no-reference view we cannot understand the generality of thought would have formed the force of such an objection, and this has been assuaged. Rather, understanding the consequences of Anscombe's argument and as such understanding self-consciousness is the remaining challenge, and one that will not be tackled in this thesis – merely revealing that such a puzzle remains is the limit of this work.

7. Bibliography

Anscombe, E. (1963) *Intention*. 2nd edn, *Intention*. 2nd edn. London: Harvard University Press.

Anscombe, E. (1981) *The Collected Papers of Elizabeth Anscombe - vol.2*. 1st edn. Oxford: Blackwell. doi: 10.1038/313163a0.

Bar-On, D. (2012) 'Expression, Truth and Reality: Some Variations on Themes from Wright', in Coliva, A. (ed.) *Mind, Meaning and Knowledge: Themes from the Philosophy of Crispin Wright*. 1st edn. Oxford: Oxford University Press, pp. 162–192.

Bilgrami, A. (2012) *Self-Knowledge and Resentment*. 1st edn. Cambridge: Harvard University Press.

Boyle, M. (2009) 'Two Kinds of Self-Knowledge', *Philosophy and Phenomenological Research*, 78(1), pp. 133–164.

Boyle, M. (2011) 'Transparent self-knowledge', *Aristotelian Society Supplementary Volume*, 85(1), pp. 223–241.

Burge, T. (1996) 'Our Entitlement to Self Knowledge I', *Proceedings of the Aristotelian Society, New Series*, 96, pp. 91–116.

Byrne, A. (2005) 'Introspection', *Philosophical Topics*, 33(1), pp. 79–104.

Byrne, A. (2011) 'I—Alex Byrne: Transparency, Belief, Intention', *Aristotelian Society Supplementary Volume*, 85(1), pp. 201–221. doi: 10.1111/j.1467-8349.2011.00203.x.

Byrne, A. (2018) *Transparency and Self-Knowledge*. 1st edn. Oxford: Oxford University Press.

Cassam, Q. (2014) *Self-Knowledge for Humans*. 1st edn. Oxford: Oxford University Press.

Castaneda, H. (1966) 'He, a Study in the Logic of Self-Consciousness', *Ratio: An International Journal of Analytic Philosophy*, 8, pp. 130–157.

Coliva, A. (2016) *The Varieties of Self-Knowledge*. 1st edn. London: Palgrave Macmillain.

Conlan, P., Merlo, G. and Wright, C. (2020) 'Eyes Directed Outward: Alex Byrne: Transparency and Self-Knowledge', *Journal of Philosophy*, 117(6), pp. 332–351.

Davidson, D. (1984) 'First Person Authority', *Dialectica*, 38(2/3), pp. 101–111.

Edgley (1969) *Reason in Theory and Practice*. 1st edn. London: Hutchison.

Evans, G. (1982) *The Varieties of Reference*. 1st edn. Edited by J. H. McDowell. New York: Oxford University Press.

Fernandez, J. (2013) *Transparent Minds: A Study of Self-Knowledge*. 1st edn. Oxford: Oxford University Press.

Finkelstein, D. H. (2012) 'From Transparency to Expressivism', in Abel, G. and Conant, J. (eds) *Rethinking Epistemology 2*. 1st edn. Berlin: De Gruyter, pp. 101–118.

Gallois, A. (1996) *The World Without, the Mind Within*. 1st edn. Cambridge: Cambridge University press.

Haddock, A. (2019) '"I am NN" : A Reconstruction of Anscombe's "The First Person"', *European Journal of Philosophy*, (December 2018), pp. 957–970. doi: 10.1111/ejop.12445.

Hampshire, S. (1975) *Freedom and the Individual*. 1st edn. Princeton: Princeton University Press.

Kant, I. (1929) *The Critique of Pure Reason*. 3rd edn. Edited by N. Kemp Smith. New York: Palgrave Macmillain.

Kaplan, D. (1989) *Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals*, *Themes from Kaplan*.

Kern, A. (2017) *Sources of Knowledge*. 1st edn. Cambridge: Harvard University Press.

Marcus, E. (2016) 'To Believe is to Know That You Believe', *Dialectica*, 70(3), pp. 375–405.

Marcus, E. and Schwenkler, J. (2018) 'Assertion and transparent self-knowledge', *Canadian Journal of Philosophy*. Routledge, 00(00), pp. 1–17. doi: 10.1080/00455091.2018.1519771.

McDowell, J. H. (1996) *Mind and World*. 1st edn. Cambridge: Harvard University Press.

Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. 1st edn. Princeton: Princeton University Press.

Nagel, T. (1986) *The View From Nowhere*. 1st edn. New York: Oxford University Press.

O'Brien, L. (2005) 'Self-Knowledge, Agency and Force', *Philosophy and Phenomenological Research*, 71(3), pp. 580–601.

Pryor, J. (1999) 'Immunity to Error through Misidentification', *Philosophical Topics*, 26(1&2), pp. 271–304.

Sellars, W. (1997) *Empiricism and the Philosophy of Mind*. 3rd edn. Edited by R. Brandom. Cambridge: Harvard University Press.

Shoemaker, S. S. (1969) 'Self-Reference and Self-Awareness', *The Journal of Philosophy*, 65(19), pp.

555–567.

Shoemaker, S. S. (1996) *The First-Person Perspective and Other Essays*. 1st edn, *Society*. 1st edn. Cam: Cambridge University press. doi: 10.2307/2653496.

Strawson, P. (1966) *The Bounds of Sense*. 1st edn. London: Methuen.

Strawson, P. (2003) *Individuals: An Essay in Descriptive Metaphysics*. 3rd edn. London: Routledge.

Williamson, T. (2000) *Knowledge and its Limits*. 1st edn. Oxford: Oxford University Press.

Williamson, T. (2007) *The Philosophy of Philosophy*. 1st edn. Oxford: Blackwell.

Wittgenstein, L. (1958) *The Blue and Brown Books*. 1st edn. Oxford: Blackwell.

Wright, C. (2001) *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*. 1st edn. Cambridge: Harvard University Press.

Wright, C. (2015) 'Self-Knowledge: the Reality of Privileged Access', in *Externalism, Self-Knowledge and Scepticism: New Essays*, pp. 49–74.