

# Reflective Equilibrium: Epistemic Desiderata, Meta-Justification, and Truth

Anders Johan Schoubye

Graduate Studies in Philosophy, University of Copenhagen  
Dept. of Philosophy, Education, and Rhetoric  
E-mail: schoubye@hum.ku.dk  
Supervisor: Associate Professor Klemens Kappel

December, 2003

## Introduction

Although John Rawls' proposed model of moral theorising, the method of reflective equilibrium, has encountered a variety of objections, many contemporary moral philosophers consider it an adequate method of moral theorising. In this paper, I intend to expound a problem intimately related to construing the method of reflective equilibrium in terms of a coherence theory of justification, as it has traditionally been. Difficulties related to the method of reflective equilibrium construed in such terms are by no means a novel or unfamiliar debate in moral epistemology and moral theory. However, one seemingly important aspect in justifying the method of reflective equilibrium has been somewhat neglected in contemporary debates. Traditionally a variety of epistemic features are usually considered inextricably connected to the meaning of coherence. Yet, the discussion of whether or not such epistemic features can be justified or even need to be justified has been somewhat overlooked. Consequently, my central contention in this paper is that these epistemic features do require an independent justification, a meta-justification, if epistemic interpretations of the method of reflective equilibrium are to succeed. Furthermore, I argue that such a meta-justification is indispensable if certain familiar objections to the method of reflective equilibrium are to be effectively refuted. Finally, I contend that the prospects of providing such a meta-justification are rather dim.

In the first section of the paper I provide a traditional outline of the method and in addition I account for some of the essential epistemic features related to the method. In the second section I account for some familiar objections to the method of reflective equilibrium. In the third and final section I discuss a variety of attempts to refute these objections. In this section I also outline my argument against these proposed defences of the method of reflective equilibrium and demonstrate why a meta-justification of the epistemic features (the epistemic desiderata) is necessary.

## The Method of Reflective Equilibrium

John Rawls' method of reflective equilibrium provided a model for moral theorising and moral justification that was to have an enormous influence on contemporary philosophy. The method of reflective equilibrium is a systematic procedure for reaching a satisfactory moral theory. The process amounts to the following:

In the first step of the process, one identifies a small class of *considered moral judgments*. These can be compared to moral intuitions or firmly held moral beliefs. What constitutes a *considered* moral judgment is a belief that seems correct under conditions usually conducive to making sound judgments. In other words, in order to identify a considered moral judgment, one should be well-informed about the issue at hand, maintain a sound inference pattern, and avoid any distorting factors such as bias, prejudice, self-interest, etc.

At the second stage of the process, one then proceeds by constructing a small set of moral principles that explain and support one's considered moral judgments. In order for these principles to function in their explanatory and supporting role, they should transcend being mere ad hoc explanations of each considered moral judgment. Instead, by identifying a general feature in the propositional content of the considered moral judgments, a moral principle should be able to account for a wide range of these judgments.

In the final step of the process, the moral principles decided upon are contrasted with the considered moral judgments. At best, the moral principles should entail all considered moral judgments and in additional cases, for instance hypothetical cases, provide judgments that also are intuitively plausible. If one encounters discrepancies (and one most certainly will), then either the principles or the considered moral judgments should be revised to reach a satisfactory match. However, if for instance a considered moral judgment is inconsistent with a moral principle, there are no clear answers as to whether one should reject particular moral principle or abandon the considered moral judgment. At this point, one must carefully evaluate how strong one's intuitions regarding the considered moral judgments are, and just how well the proposed moral principle succeeds in explaining additional considered moral judgments.

Although the method of reflective equilibrium has been outlined in various ways by numerous moral epistemologists, the previous rough sketch accords fairly well with traditional accounts of the method.<sup>1</sup>

It is rather evident that the method of reflective equilibrium shares a significant amount of similarities with the coherence theory of justification and the method has traditionally been interpreted in such terms. As a result, more elaborate expositions of the method of reflective equilibrium have often attempted to describe the characteristic features of coherent beliefs and coherent believing. In the following I refer to these characteristic features as *epistemic desiderata*. A rough and rather vague sketch of such epistemic desiderata might include the following:

---

<sup>1</sup> For other rough sketches cf. e.g. Scanlon, Daniels, Sayre-McCord, Kagan, Brink, and Dworkin.

**Consistency:** Contradictions must be avoided in the system of beliefs to maintain coherence. No system that contains both the belief that  $p$  and the belief that  $\neg p$  can be considered coherent.

**Simplicity:** All things being equal, a simple set of basic moral principles is preferred to a complex set of basic moral principles.

**Power:** All things being equal, the more plausible beliefs a set of basic moral principles can justify – the more powerful the system.

**Generality:** All things being equal, the more general the beliefs, the better the system. A set of principles should encompass as large an area as possible.<sup>2</sup>

**Systematicity:** Explanatory relations should obtain among beliefs in the particular system of beliefs. The explanatory relations should account for the morally relevant factors of a given moral judgment and thus adequately account for morally relevant distinctions<sup>3</sup>

**Intuitive plausibility:** A set of moral beliefs should fit our considered moral judgments.

Of course, the above mentioned features do in no way exhaust the notion of epistemic desiderata. It is possible and even very likely that further epistemic features could and should be added to the list. Furthermore, the list is by no means a precise or accurate definition of each epistemic desideratum, but merely an approximate description.

In simplified terms, a moral theory can be considered satisfactory to the extent it produces plausible explanations. But what constitutes a plausible explanation will depend largely on the structure of the theory. If moral principles are to provide an explanation and hence a justification of particular beliefs, then some levels of the epistemic desiderata must be satisfied. If this seems unreasonable, consider a moral theory comprised of ad hoc moral principles arranged simply to comply with some individual's idiosyncratic view on morality. No one would ever accept such a theory, because in regards to moral theorising we expect some levels of generality, simplicity, systematicity etc. to be included. Ad hoc moral principles seem entirely devoid of explanatory power and provide no satisfactory justification of particular moral beliefs. It is quite evident that a moral theory as the above that completely disregards epistemic features would normally be considered entirely unjustified and outright false. Adequate and intuitively plausible justification seemingly necessitates the satisfaction of epistemic desiderata.

However, I do not intend to suggest that a moral theory should satisfy all levels of epistemic desiderata to a certain degree or even to an equal degree. Instead, trade-offs are usually considered permissible to increase overall coherence. One could add a moral

---

<sup>2</sup> Klemens Kappel pointed out to me that generality and power in essence are identical features. However, I regard power as the amount of plausible answers that a specific principle can yield in all cases, whereas by generality, I mean a principle encompassing larger areas. Yet, a general principle that provides plausible moral judgments regarding, for instance, nature or animals may simply be said to be a more powerful principle. However, a powerful principle may just yield more answers in one aspect of morality without thereby being a more general principle. So in certain respects, generality and power are separable.

<sup>3</sup> cf. Kagan's 'dangling distinctions in 'The Limits of Morality' – p.14.

principle that decreases, say, the overall simplicity of one's theory, if it, for instance, adds substantially to the theory's overall systematicity – and thereby increases overall coherence. As such, trading-off on one level may be entirely reasonable insofar as the theory gains at some other level and thus increases in overall coherence. Yet the exception to this rule is of course consistency. While trading off on the other levels may be entirely justified, trading of consistency would entail false beliefs in the system. The belief that  $p$  and the belief that  $\neg p$  cannot both be true, and therefore consistency must be maintained.

On some interpretations of the method of reflective equilibrium, the epistemic desiderata are considered conducive to truth. On such interpretations, the success of a moral theory depends largely on how well the theory *satisfies* the epistemic desiderata. Satisfaction is to be understood as the extent to which the theory includes levels of simplicity, generality, power, systematicity. One could say that a system of moral beliefs is considered coherent to the degree it contains explanatory and justificatory force. In this respect, accepting some levels of simplicity, power, generality etc. in the theory plays an essential role. The central contention in this paper is that it remains unclear why including levels of simplicity, generality, power etc. renders the method conducive to truth. I argue that in order to justify such an assumption a meta-justification of the epistemic desiderata is necessary.

However, I will not discuss additional familiar objections to the method of reflective equilibrium. I presuppose that considered moral judgments could constitute a satisfactory basis for truth-conducive moral theory-building and moreover that the method of reflective equilibrium avoids charges of being intuitionism in disguise.<sup>4</sup> Neither will I discuss foundationalist interpretations of the method of reflective equilibrium. Such interpretations might also encounter a need for meta-justifying epistemic desiderata, but I focus exclusively on suggestions that have interpreted the method in terms of a coherence theory of justification.

## Objections to the Method of Reflective Equilibrium

Traditionally, two chief objections to the method of reflective equilibrium have been meticulously rehearsed, both of which are variants of familiar objections to the coherence theory of justification. These objections are somewhat closely related and do in essence stress a quite similar difficulty in the method of reflective equilibrium.

### The Truth Objection<sup>5</sup>

Utilising the method of reflective equilibrium should ideally yield a maximally coherent set of considered moral judgments as well as moral principles – such that the moral principles adequately explain the considered moral judgments. However, despite maximal coherence, it is quite possible to imagine that such a system of beliefs could be entirely untrue:

---

<sup>4</sup> For additional discussions of the objection that considered moral judgments cannot constitute a satisfactory basis for moral theorising cf. e.g. R.M. Hare. Elaborate discussions of reflective equilibrium being intuitionist cf. e.g. Norman Daniels, T.M. Scanlon, and David Lyons.

<sup>5</sup> The following section owes much of its merit to Tersman's account of the truth objection in 'Reflective Equilibrium – an Essay in Moral Realism' as well as Klemens Kappel's paper 'Is there a Meta-justification of Reflective Equilibrium?'

Even though a system of beliefs may enjoy maximal coherence, it remains quite possible that the beliefs in that particular system nonetheless turn out to be false.

This is in essence what the truth objection states and the claim seems quite conceivable; just imagine various sceptical scenarios such as Cartesian demons or brains in vats. The challenge for a proponent of the method of reflective equilibrium is to explicate why coherence leads to true beliefs.

The truth objection derives its content from a similar objection to the coherence theory of justification traditionally referred to as the *isolation problem*<sup>6</sup> and it has been stated in various ways by a wide variety of critics. However, despite different phrasings and diverse points of emphasis, it seems that the general idea of the objections have pointed towards a distinct problem of truth. For instance, consider David Lyons' charge against the method of reflective equilibrium,

(..) the justificatory force of coherence arguments is unclear. Suppose one assumes that there are such things as valid principles of Justice which can be justified in some way; suppose one believes, moreover, that a coherence argument explicates our shared sense of justice, giving precise expression to our basic moral convictions: one may still doubt whether a coherence argument says anything about the validity of such principles. (Lyons: 1976 p.147)

Although Lyons avoids explicitly mentioning truth, his contention is certainly a variety of the truth objection: it remains uncertain whether the principles decided upon through the method of reflective equilibrium are *valid*, or, in other words, there are no means of establishing whether or not these principles are in fact *true*. Even though the principles may yield an explanation of our considered moral judgments and may moreover be maximally coherent, the question remains whether these facts alone guarantee truth.

However, in reply to the truth objection some theorists have argued that we should be fallibilists regarding justification. That is, justification should not be taken as guaranteeing truth. Given the assumption that a wide variety of beliefs that are considered constitutive of knowledge *could* turn out false<sup>7</sup> – such demanding criteria for justification should be rejected.<sup>8</sup> Such theorists therefore maintain that while coherence as such cannot be said to guarantee truth, it can be regarded a truth-conducive method of justification. In other words, if one systematically ensures that one's system of beliefs is coherent, then the likelihood of one's beliefs being true increases. Accordingly, coherence can be said to provide true beliefs in a majority of cases, while maintaining the possibility that coherence could also yield false beliefs.

---

<sup>6</sup> Bonjour: The Dialectic of Foundationalism and Coherentism p. 127.

<sup>7</sup> Only a wide variety, since some reservations should be preserved in regards to for instance a priori beliefs and necessarily true beliefs.

<sup>8</sup> cf. David Brink: Moral Realism and the Foundations of Ethics – p. 126.

## The Plurality Objection

The second objection to the method of reflective equilibrium relies on another objection to the coherence theory of justification on occasion referred to as the *alternative coherent systems objection*.<sup>9</sup> In regards to the method of reflective equilibrium, it has been referred to as the *plurality objection* or *the implausible relativity objection*.<sup>10</sup> In short, the objection states the following:

It is quite possible that there could be several radically different and incompatible systems of beliefs that are equally coherent and thus equally justified. However, since only one coherent system of beliefs can be true, there is no obvious reason to think that coherence is truth-conducive.

On one interpretation, the plurality objection touches on the same problem as that of the truth objection. However, on an extended interpretation of the plurality objection, it also states that there are no obvious reasons for thinking that coherence as such is even truth-conducive. Insofar as several incompatible yet maximally coherent systems of beliefs might exist, why assume that coherence even in general leads to truth? Why not instead assume that coherence *leads away from truth rather than towards it*?<sup>11</sup>

Some critics have therefore maintained that in order for the method of reflective equilibrium to be truth-conducive, some of the primary beliefs in the belief system must enjoy an initial credibility. This is most notably explicated in an often cited passage from Richard Brandt's 'A Theory of the Good and the Right',

There is a problem here which is quite similar to that which faces the traditional coherence theory of justification of belief: that the theory claims that a more coherent system of beliefs is better justified than a less coherent one, but there is no reason to think that this claim is true unless some of the beliefs are initially credible – for some reason other than coherence, say because they state facts of observation (Brandt: 1979 – A Theory of the Good and the Right)

However, I find it somewhat unclear what Brandt more precisely has in mind by *credibility*. One could argue that Brandt has epistemically privileged beliefs in mind, such as *foundational* beliefs.<sup>12</sup> But then it seems that Brandt is merely insisting on a foundationalist theory of justification, rather than stating a genuine objection. Instead, I think Brandt's objection should be construed as pointing out that coherence as such cannot be *truth-conducive* unless some of the primary beliefs in the belief system are *true*.<sup>13</sup> For instance, if

---

<sup>9</sup> cf. for instance BonJour: The Dialectic of Foundationalism and Coherentism – p. 128.

<sup>10</sup> cf. T.M. Scanlon: Rawls on Justification – p.151.

<sup>11</sup> Richard Brandt: A Theory of the Good and the Right p. 20.

<sup>12</sup> Foundationalist theories of justification maintain that there is a class of epistemically privileged beliefs that are self-justified or non-inferentially justified. Such foundational beliefs mean to avoid the sceptic objection of an infinite regress. However, I will not elaborate any further on foundationalist theories of justification.

<sup>13</sup> On this interpretation, coherence is construed as the main reason for regarding something likely to be true. It is important to notice that this is merely an interpretation of Brandt's objection, since it is possible to support a coherentist theory of justification that does not support the claim that coherence as such is truth-

a system of beliefs is founded on a small set of considered moral judgments that are entirely false, then, albeit admitting maximal coherence, it seems quite implausible that adding further cohering beliefs to the system will increase its likelihood of being true. It seems that there is no obvious reason to regard coherence as such as truth-conducive.

It is possible that an adequate reply to these objections could be provided without relying on the before mentioned epistemic desiderata. But it seems to me that some of the most promising replies to these objections do in fact implicitly or explicitly rely on these epistemic features in their arguments. In the following I intend to demonstrate why these replies necessitate a meta-justification of the epistemic desiderata in order to succeed. However, before I proceed to these replies, I want to remark briefly on non-epistemic interpretations of the method of reflective equilibrium

### **Additional Interpretations of the Method of Reflective Equilibrium**

There is one important point regarding the objections that I have accounted for in the previous section that needs some emphasis: the difficulties explicated above pertain predominantly to *epistemic* interpretations of the method of reflective equilibrium. By epistemic interpretations, I mean interpretations that aim to link the method of reflective equilibrium with some notion of truth. Moreover, even though epistemic interpretations do not require embracing any particular notion of truth, they are quite often combined with a defence of moral realism. And, combining moral realism with the method of reflective equilibrium obliges one to answer certain rather pressing questions such as: in what way can a reflective equilibrium be said to epitomize a moral reality? And why assume that the method is truth-conducive or provides a true moral theory?

However, it should be noted, that there is a wide variety of additional interpretations of the method, as for instance *moral*, *pragmatic*, and *descriptive* interpretations, that remain in opposition to the epistemic interpretation. These alternative interpretations suggest quite different purposes for using the method of reflective equilibrium.

A *moral* interpretation of the method might maintain that we are morally required to act on the basis of moral principles, and consequently that we should seek to systematize our moral principles to gain consistency in our moral judgments. A *pragmatic* interpretation might suggest that it is useful if our moral principles are generated on the basis of some systematic account. Geoffrey Sayre-McCord offers one such possible interpretation,<sup>14</sup>

Yet another way to defend the method, again without appealing to its epistemic value, is to argue that its use has significant practical advantages. Just as all sorts of advantages are secured by requiring that judges make explicit the rationale behind their decisions, so too, we might think, practical considerations require something similar of ordinary people. After all, moral thought and reflection obviously play a crucially important role in social life; to the extent this role might best be served by people being able to articulate and defend the principles on which they act, using and recommending the method of reflective equilibrium would seem eminently reasonable. (Sayre-McCord 1996: p.10)

---

conducive. See for instance Sayre-McCord, who provides such a version of the coherence theory of justification.

<sup>14</sup> However, Sayre-McCord does in fact support an epistemic interpretation.

A *descriptive* interpretation might state that the method of reflective equilibrium should be construed as a method for determining the structure of an individual's (or group's) moral beliefs and moral conceptions. For instance, in identifying the considered moral judgments held true by a particular individual, then by the method of reflective equilibrium we can describe a set of moral principles, which that person would accordingly hold true. On the descriptive interpretation, the method of reflective equilibrium is thus utilised to describe the moral theory supported by particular individuals (or groups).<sup>15</sup>

Even though the method of reflective equilibrium has often been conceived of in epistemic terms, these non-epistemic interpretations deserve mentioning, as they to some extent avoid the difficulties stated in the truth and plurality objections. Given that they have no intention of combining the method with any notion of truth, it should be clear why this is the case.

Finally, some remarks on combining the method of reflective equilibrium with a constructivist notion of truth. One such interpretation, indeed a particularly fervent one, is expounded by Ronald Dworkin.<sup>16</sup> On Dworkin's account, the method of reflective equilibrium is construed as a method for *constructing* adequate moral principles, as opposed to *discovering* them.<sup>17</sup> Dworkin rejects moral realism and the idea that a reflective equilibrium epitomizes a moral reality. Instead constructivists such as Dworkin maintain that the method of reflective equilibrium is an instrumental device for establishing acceptable moral rules.<sup>18</sup> These moral rules serve as action-guiding measures and ensure that our moral judgments are consistent and coherent. Thus, the constructivist approach links the method of reflective equilibrium to practical reasoning instead of theoretical reasoning. Moral principles cannot be true in epistemic terms, but they can be valid in the sense that people accept these principles as fair and just, and act in accordance with them. The method of reflective equilibrium is designed for constructing such principles.

The requirement of a meta-justification of the epistemic desiderata pertains predominantly to theories that defend moral realism and an epistemic interpretation of the method of reflective equilibrium.<sup>19</sup> In the following sections I shall offer an outline of three very different suggestions that contest the truth and plurality objections; these suggestions are all characterized by defending moral realism and interpreting the method of reflective equilibrium in epistemic terms.

---

<sup>15</sup> cf. T.M. Scanlon's: Rawls on Justification – p.142, for an outline of such an interpretation.

<sup>16</sup> cf. Ronald Dworkin: The Original Position – p.28ff in 'Reading Rawls' ed. Norman Daniels.

<sup>17</sup> cf. Dworkin: The Original Position – p.27 in 'Reading Rawls' ed. Norman Daniels.

<sup>18</sup> In many ways, the constructivist approach and the pragmatic interpretation mentioned earlier have certain similarities. But the constructivist approach builds in the moral requirement that we have a responsibility to seek coherent and consistent moral principles in order to exercise moral judgments. However, one could perhaps argue that such a requirement is implicit in the pragmatic interpretation.

<sup>19</sup> Although I find it quite possible that constructivist interpretations of the method might encounter similar problems.



## Replies to the Truth and Plurality Objections

### David Brink & Reliable Moral Beliefs

A somewhat straightforward defence of the method of reflective equilibrium is provided in ‘Moral Realism and the Foundations of Ethics’ by David Brink. Brink defends an epistemic interpretation of the method and argues that notwithstanding the seemingly strong objections, a coherence theory of moral justification is compatible with moral realism. In the following section I present a fairly simplified exposition of Brink’s reasoning. Subsequently, I offer an objection to Brink’s defence of the method of reflective equilibrium.

### First-order & Second-order Beliefs

A coherence theory of *moral* justification should according to Brink identify *theoretical moral claims* or, in other words, a set of moral principles that provide and explain moral judgments.

If the method of reflective equilibrium is to be truth-conducive, evaluating the proposed theoretical claims (the moral principles) requires that they are contrasted with moral beliefs that are *reliable*. Brink argues that to identify *reliable* moral beliefs certain epistemic conditions regarding one’s moral beliefs must be taken into account. He therefore introduces a notion of *second-order* beliefs.<sup>20</sup> Second-order beliefs are beliefs that concern relevant epistemic conditions for forming beliefs that are likely to be true. That is, they are beliefs concerning what constitutes a reliable belief. Consequently, other theorists have referred to second-order beliefs as *epistemic beliefs*.<sup>21</sup>

Generally speaking, a first-order belief *p* has a propositional content expressing some fact or state of the world (e.g. the belief that grass is green). Conversely, a second-order belief may possibly concern what kind of belief *p* is, explain why such a *p*-type belief is likely to be true, or possibly describe the circumstances under which *p* was formed. A first-order belief is considered reliable if it by and large satisfies the epistemic conditions considered essential to truth-conducive belief-forming. For example, if I have the first-order belief *q* that I am a 25-year-old man, then I should also have a variety of epistemic beliefs regarding the reliability of my memory, my capability of counting, whether or not I am under the influence of memory-distorting drugs etc. These epistemic conditions pertain to the reliability of my first-order belief *q*.

Brink rejects the demand for infallibilism regarding justification, which he refers to as ‘objectivism about truth’. He argues that only deductive reasoning will guarantee true beliefs and that a wide range of beliefs usually considered constitutive of knowledge could in fact turn out false.<sup>22</sup> Thus, since a wide range of non-deductive beliefs are considered constitutive of knowledge, the requirement that justification must provide true beliefs is

---

<sup>20</sup> In regard to second-order beliefs, Brink relies solidly on the coherence theory of justification as developed by Laurence Bonjour in ‘The Structure of Empirical Knowledge’ (1985).

<sup>21</sup> cf. Michael Williams: Unnatural Doubts – p. 284.

<sup>22</sup> David Brink: Moral Realism and the Foundations of Ethics – p.126.

much too strong. Epistemologists should instead, according to Brink, embrace the notion of second-order beliefs, as these initially provide an explanation of why we can consider certain beliefs to be reliable. Furthermore, if second-order beliefs maintain internal coherence and additionally cohere with the first-order beliefs, one then has the *best possible* reason for regarding one's beliefs as being likely to be true. If first-order beliefs regarding the external world, our surrounding environment, etc. – and second-order beliefs regarding relevant epistemic conditions (such as our *psychological makeup*, *perceptual equipment* etc.<sup>23</sup>) cohere, the overall belief system provides the best overall explanation of an external world that we reliably detect and interact with.

Since the realist second-order beliefs in question are, ex hypothesi, maximally coherent, there is the best kind of reason we can have for holding them to be true. Our assumption is that certain realist second-order beliefs provide the best explanatory account of a wide range of observational beliefs consistently with other theoretical beliefs we hold. Such beliefs might turn out to be false, but it is surely neither crazy nor unreasonable for us to hold them. (Brink: 1989 p.129).

Thus, second-order beliefs outline epistemic conditions that must be satisfied to form reliable beliefs.

With regards to moral theorising, Brink states that this is merely a special case of first-order and second-order beliefs. In moral theory-building first-order beliefs are *moral* beliefs concerning for instance what one considers right or wrong actions and which moral principles one holds true. Conversely, second-order beliefs are *non-moral* beliefs concerning such things as the nature of morality or the relevant epistemic conditions for generating reliable *moral* beliefs. Such second-order beliefs could include beliefs in moral realism or beliefs pertaining to the relation between moral beliefs and the world. Some of the epistemic conditions that moral beliefs should satisfy are in fact those mentioned in the outline of the method of reflective equilibrium.

A belief that is based on the available (nonmoral) evidence and is thus well informed, that results from good inference patterns, that is not distorted by obvious forms of prejudice or self-interest, that is held with some confidence, and that is relatively stable over time is formed under conditions conducive to truth. These conditions of general cognitive reliability confer some reliability on moral beliefs so formed (Brink: 1989 p.132).

However, moral beliefs can gain additional reliability. One might presume that there are morally motivated epistemic conditions that should also be satisfied in order to form reliable moral beliefs. Consequently, one could have additional second-order beliefs containing such morally motivated epistemic conditions. Nonetheless, morally motivated conditions should encompass such a generality that they remain morally motivated epistemic conditions regardless of the moral theory that one possibly accepts.<sup>24</sup>

---

<sup>23</sup> cf. Brink: Moral Realism and the Foundations of Ethics – p.127.

<sup>24</sup> cf. Brink: Moral Realism and the Foundation of Ethics - p.132. To illustrate, a morally motivated condition stating that a moral belief is reliable only insofar as it reduces morality to a question of maximizing overall utility would not be acceptable.

Brink states that one can plausibly defend moral realism to the extent one accepts justification as involving second-order beliefs. In fact, Brink maintains that second-order beliefs are *realist* second-order beliefs, given that they concern our relation to an external world that is *metaphysically or conceptually independent of our evidence about it*.<sup>25</sup> The second-order beliefs are realist second-order beliefs because they outline the epistemic demands that must be satisfied in order to form reliable moral beliefs.

### **Brink's Response to the Plurality Objection**

Brink argues, in response to Brandt's objection, that considered moral judgments *do* have initial credibility - even prior to reaching a reflective equilibrium. This initial credibility is constituted by the fact that the considered moral judgments are the result of a reliable belief-forming mechanism delineated by second-order beliefs. Brink writes in reply,

[...] the credibility of considered moral beliefs can be established by appeal to psychological and moral theories before determining which moral beliefs are part of a maximally coherent system of beliefs. Thus, considered moral beliefs do have initial credibility and are not just initially believed (Brink: 1988 p.136).

However, Brink acknowledges that only the end result of a reflective equilibrium will provide a moral theory that is likely to be true. A system of beliefs is only *systematically* justified to the extent that one's entire system of beliefs is coherent (here, including second-order beliefs at various levels).<sup>26</sup> However, initially one can regard a moral belief as reliable (*contextually* justified) insofar as one satisfies the epistemic conditions outlined by one's second-order beliefs.

Brink's compares his claim to the status of observational beliefs in scientific theory-building. He emphasises that even in natural sciences theory acceptance necessitates a coherence theory of justification; observational beliefs in isolation lose their value too, as observational beliefs are theory dependent and must cohere with theoretical hypothesis. Consequently, justification requires second-order beliefs concerning the reliability of observational beliefs as well as coherence with additional scientific theories. Brink, in other words, argues that truth-conducive theory-building requires second-order beliefs both in the natural sciences as well as in ethics.

In moral philosophy we want to apply the same sort of criteria that we use for theory building quite generally: we want our moral theory to have simplicity, power and coherence (Kagan: 1988, p.11).

---

<sup>25</sup> Brink: Moral Realism and the Foundation of Ethics - p.127.

<sup>26</sup> Although I cannot elaborate fully on the concepts of *systematic justification* and *contextual justification* here, the following outline should suffice: *Contextual justification* can be understood as a type of initial justification, where certain supporting premises are presupposed (here second-order beliefs play a vital role). *Systematic justification* is to be understood as the categorical justification of all beliefs and premises in a system of beliefs.

## Epistemic Desiderata & The Plurality Objection

It seems quite reasonable to assume that Brink would concede that epistemic justification involves the satisfaction of epistemic desiderata such as those mentioned in the first section of this paper. According to Brink, explanatory power remains the best possible reason for assuming that the method of reflective equilibrium is truth-conducive. Certainly he agrees that in order to provide such explanatory power the theoretical claims must include some levels of simplicity, generality, power etc.

[...] explanatory coherence demands that we introduce more general, theoretical moral claims into our moral views in order to extend our moral views to new cases, [...] to try to resolve disagreements with others over particular moral issues, and to unify and explain the more particular moral views we already hold (Brink: 1989 – p.130).

However, some difficulties in Brink's account of the method of reflective equilibrium emerge if one considers a slightly modified version of the plurality objection.

For example: say we face two different systems of beliefs  $S_1$  and  $S_2$ . We presuppose that both systems are consistent and that the beliefs justified by each system are intuitively plausible. However, imagine that  $S_1$  is a more simple system than  $S_2$ :  $S_1$  justifies as wide a range of moral beliefs as  $S_2$ , but from a lesser amount of moral principles. Which of these systems would Brink consider more likely to be true?

I think Brink would argue, quite reasonably, that system  $S_1$  has more explanatory power than  $S_2$ , because  $S_1$  needs fewer principles to explain as wide a range of beliefs as  $S_2$ . Since it is the explanatory power of coherence that, according to Brink, is the *best possible*<sup>27</sup> reason for regarding coherence conducive to truth,  $S_1$  is more likely to be true than  $S_2$ .

Yet, and this is my objection to Brink's reasoning, if both  $S_1$  and  $S_2$  succeed in explaining our reliable moral beliefs, it seems to me that there is no obvious reason for thinking that settling for the simple system will increase the likelihood of that particular system being true? For all we know, a complex system might just as well be the *true* set of moral principles. Thus, in regards to epistemic justification one seems entirely as justified in settling for either set, regardless of its complexity.

Consider another example: again we face two different systems of beliefs  $S_1$  and  $S_2$  where it is presupposed, as earlier, that both systems are consistent and that the beliefs justified by each system are intuitively plausible.

However, imagine that while  $S_1$  is a very powerful but also very complicated system of beliefs,  $S_2$  is extremely simple but not quite as powerful as  $S_1$ . In other words, while  $S_1$  justifies a large number of moral beliefs, it also contains a wide range of moral principles. Conversely,  $S_2$  justifies a small amount of moral beliefs, but also contains very few moral principles. Now, which of these systems would Brink consider more likely to be true? Due to his emphasis on explanatory power, I think Brink would choose  $S_2$ . Provided that  $S_2$  is more powerful than  $S_1$  and thus justifies more plausible moral beliefs than  $S_1$ ,  $S_2$  is more explanatory powerful.

---

<sup>27</sup> cf. Brink: Moral Realism and the Foundations of Ethics – p.129.

However, one could then reasonably ask: why not settle for an even more complicated system of beliefs? A system in which every possible considered moral judgment is justified by ad hoc principles assigned to each belief?

In my opinion, the only plausible reply to this question is the following: to the extent that moral judgments are justified merely by ad hoc principles, principles become somewhat redundant. Moral principles (theoretical claims) are needed in order to *explain* and *thus* justify our considered moral judgments. But in the case of ad hoc principles, the principles provide no explanations and the moral judgments yielded by such principles seem entirely unjustified. It seems that moral principles only provide an explanation of our moral judgments if they encompass a somewhat wide variety of considered moral judgments. In other words, there are limits as to how complex a theory can be without trading-off intuitive plausibility.

In the above examples an appeal to second-order beliefs will not work, because it is presupposed that the epistemic conditions pertaining to forming reliable beliefs are already satisfied.

In my view, there are good reasons for assuming that the epistemic desiderata are conducive to truth. Certainly, a theory that does not satisfy some levels of simplicity, generality, power, etc. will lack explanatory power and as such justificatory force. However, the difficulty is that in the above examples it is entirely unclear which system is more likely to be true. This is so because there are no means of deciding to what extent these epistemic features must be satisfied. Consequently, I think that a meta-justification of the epistemic desiderata is needed – a justification that accurately explains why these epistemic features are conducive to truth. Moreover, I think that an explication of the ‘proportionate value’<sup>28</sup> of each epistemic desideratum is also needed. This is the case because a) without a meta-justification of the epistemic desiderata it cannot be demonstrated why a simple system is, *ceteris paribus*, supposedly better than a complex one. And b) without some idea of the proportionate value of each desideratum (to what extent they must be satisfied), it cannot be determined whether or not a trade-off between the various desiderata is reasonable. In conclusion, I think that the plurality objection in the above modified sense cannot be rejected, unless such a meta-justification is provided.

### **Folke Tersman – Coherence & the Principle of Charity**

Another attempt to challenge the truth and the plurality objections is presented in ‘Reflective Equilibrium, an Essay in Moral Epistemology’ (1993) by Swedish philosopher Folke Tersman.

Tersman’s main contention is that the plurality objection is misguided, regardless of how it is interpreted. If the plurality objection aims at demonstrating that two persons may be equally justified in holding incompatible beliefs, say the belief that  $p$  and the belief that  $\neg p$ , then, according to Tersman, it is a trivial observation. Such particular yet minor inconsistencies between different systems of beliefs may simply be explained in terms of available evidence, epistemic perspective etc. For instance, a person A may be justified in believing  $p$  given the evidence available to him, while another person B may be justified in

---

<sup>28</sup> ‘Proportionate value’ is perhaps not the most suitable wording. I mean an explication of the conditions that must obtain to make a trade-off between desiderata reasonable. That is, under what conditions is it reasonable to trade-off simplicity for power, generality for systematicity, etc.

believing  $\neg p$  given the evidence available to him. Although both systems of beliefs cannot be true, both A and B may still be justified in holding either belief to the extent that it coheres with their additional beliefs. Such minor inconsistencies cannot provide the conclusion that coherence as such is not truth-conducive.

However, if the plurality objection aims at demonstrating that *radically* different and yet incompatible systems of beliefs may be equally justified, then it does constitute a significant problem. If this is in fact possible, it would suggest that coherence as such is not truth-conducive. It is primarily as a response to this interpretation of the plurality objection that Tersman develops his argument.

Firstly Tersman argues that if a belief  $p$  coheres with person A's system of beliefs, it seems reasonable to assume that  $\neg p$  does not cohere equally well with A's system. Additionally it seems reasonable to assume that  $\neg p$  would also weaken the explanatory power of A's belief system. So, if  $\neg p$  were to fit in A's system of beliefs, then some changes (perhaps even radical changes) would have to be made in order to make  $\neg p$  cohere as well as  $p$ . The number of additional beliefs that must be changed to make  $\neg p$  cohere depends on how well  $p$  coheres with A's system of beliefs. As such, Tersman claims that the better  $p$  coheres with A's system of beliefs, the less likely it is that  $\neg p$  coheres equally well with another person's system. And consequently, the better  $p$  coheres with A's system the less likely it is that some other person actually believes that  $\neg p$ . The inference from one system of beliefs (A's) to another person's system of beliefs will become clearer in the following.

Yet, Tersman acknowledges that it is *possible* in the logical sense that two persons could be equally justified in holding radically different systems of beliefs. For instance, if we consider possible worlds, this would certainly provide thoroughly justified systems of beliefs that are nonetheless entirely false. However, admitting logical possibility, Tersman still argues that such scenarios are not possible in practice, or they are at least extremely implausible.

Tersman relies quite heavily on the notions of *radical interpretation* and *the principle of charity* as developed by Donald Davidson. To give a clear exposition of Tersman's arguments, these notions need to be explicated.

In Donald Davidson's theory of meaning the notion of radical interpretation plays an essential role. Radical interpretation is, in simple terms, Davidson's empirical method of establishing the truth value of unknown sentences. Davidson suggests that by using radical interpretation, it is possible to ascertain the meaning of an unknown sentence without knowing the speaker's beliefs or understanding the meaning of the words in the sentence.

Imagine that a speaker utters a sentence in an entirely unknown language and that we have no knowledge of the speaker's beliefs. Then by conducting radical interpretation, we embark by analysing the circumstances under which the sentence was uttered and then assign a preliminary truth value to the sentence. Say, if person A utters the sentence  $S_1$  in language L only when a cat crosses the road – we fix a preliminary truth value for  $S_1$ : for instance ' $S_1$ ' is true in L if and only if a cat crossed the road. According to Davidson, if we know the truth value of a sentence, we then know the conditions that must obtain to make the sentence true. In other words, to know the truth value of a sentence is to understand the meaning of that sentence. We then proceed to interpret all of A's sentences by affixing truth

values. If inconsistencies occur during this process, we must then revise proposed truth values of A's sentences. Yet, over time it will be possible to extract word-meaning from the more or less accurate truth values that we have instantiated. In the long run we should be able to fully understand the sentences and consequently the language spoken by A.<sup>29</sup>

If this view seems implausible, consider its negation – it implies e.g. that a child (who in learning her first language might well be described as a radical interpreter) could grow up with completely erroneous views about the meanings of the utterances of her parents (in spite of being able to interact perfectly with them on the basis of her interpretations of their utterances) (Tersman: p.106).

The principle of charity is then introduced to avoid a variety of difficulties pertaining to radical interpretation. One common objection to radical interpretation is exposed in the following observation: if A says 'a cat crossed the road', how do we know that he does not actually *mean* a cat crossed the road and  $2+2 = 4$ ? Since the sentence is true in all instances where a cat in fact crossed the road and false in all instances where a cat in fact did not cross the road, how can this be ruled out? Davidson responds that we should apply a principle of charity to the method of radical interpretation. The principle of charity states that we should not ascribe beliefs to a person, which it would be irrational for that person to hold. For instance, in the case of person A uttering 'a cat crossed the road', then applying the principle of charity demands that we do not ascribe additional meaning to the utterance that would be superfluous in the particular context. We should interpret the utterances of others such that the sentences that they hold true conform somewhat to the sentences that we hold true.

What matters is this: if all we know is what sentences a speaker holds true, and we cannot assume that his language is our own, then we cannot take even a first step towards interpretation without knowing or assuming a great deal about the speaker's beliefs. Since knowledge of beliefs comes only with the ability to interpret words, the only possibility at the start is to assume general agreement on beliefs (Davidson: 1974 – On the Very Idea of a Conceptual Scheme, Inquiries p. 196).

In other words, we should by and large regard the speaker (being interpreted) as holding predominantly true beliefs.

However, a central presumption that underlies both Davidson's theory of meaning, the notion of radical interpretation, and the principle of charity is a holistic view of meaning. Such a view suggests that the meaning of any linguistic expression or sentence in a language is dependent on relations to other linguistic expressions or sentences in the particular language. That is, if the meaning of one sentence in a language is to be established, then the meanings of a very wide range of additional sentences must already be known.

Suppose that a person accepts the sentence "most bachelors are unhappy", and we wonder if it is reasonable to attribute the belief to her that most bachelors are unhappy. What assumptions would we have to make about her beliefs? [...] it is clear that we would have to presuppose that she agrees with

---

<sup>29</sup> This is of course an overly simplified account of Davidson's theory, but it will suffice for the present purpose.

us on a vast number of issues concerning bachelors and unhappiness (Tersman: 1993, p.109).

This view is perhaps most notably expounded by Quine, who claims that any translation of an unknown language will be indeterminate. In very simple terms, Quine argues that for any translation of an unknown language there will be additional incompatible, yet entirely adequate translations available. As such, it remains indeterminate which of the available translations are correct.<sup>30</sup> Davidson's notion of radical interpretation intends to circumvent this problem. While Davidson embraces the holistic view of meaning, he nonetheless claims that by applying the principle of charity it is possible to interpret a speaker's sentences. But, maintaining the holistic view of meaning entails that a radical interpretation will succeed only if the speaker holds a wide range of additional beliefs that by and large correspond to the interpreter's. Now leaving this small excursion on Davidson, we can return to Tersman's argument.

If the notions of radical interpretation and the principle of charity are accepted, it becomes clear why Tersman regards the plurality objection as untenable.

Tersman's argument should be understood in conjunction with his initial claim:

the assumption that the better a belief  $p$  coheres with a specific system of beliefs, the worse  $\neg p$  will cohere with that particular system.

Now, imagine two different persons each having a system of beliefs that are equally coherent. Furthermore, imagine that the systems are largely in agreement, but diverge on a small amount of incompatible beliefs. Given that they already share a wide variety of beliefs, it will be possible for each of them to determine which beliefs they disagree on (and possibly why they disagree). However, if we imagine these two persons having radically different systems of beliefs, then, on Tersman's account, it becomes almost impossible for either person to determine how their beliefs diverge from each other. It is simply not possible for either person to interpret the other, because an interpretation requires that the person being interpreted by and large has beliefs that correspond to the interpreter's. Furthermore, it also becomes impossible for either person to establish whether or not the other person's system of beliefs coheres, since one must know the content of the beliefs in a system in order to determine whether the system in fact coheres. Consequently, claiming that two persons may be equally justified in holding radically different systems of beliefs disregards the principle of charity. It does so because considering such a possibility requires that we hypothesise one person as having a wide range of beliefs, which it, on our view, would be entirely irrational to hold. Further, to hypothesise a person holding such radically different beliefs, it is also presumed that it is possible to establish what the contents of those beliefs are (that we are able to interpret that person). Yet, according to Tersman, this remains highly unlikely. Why think that it can be adequately established that someone in fact holds such a radically different system of beliefs? Conceding the logical possibility of equally justified, yet radically different systems of beliefs, in practice this remains an extremely implausible possibility.

---

<sup>30</sup> cf. Quine: *Word and Object* – 1960. Again, this is an extremely simplified account of Quine's reasoning, but should suffice for present purposes.



### **Tersman & Epistemic Desiderata**

Seeing as Tersman does not address the question of epistemic desiderata, it is difficult to determine whether or not he would agree with the claim that epistemic desiderata are truth-conducive. However, I find it quite reasonable to assume that Tersman would at least agree that the satisfaction of epistemic desiderata increases overall coherence. For instance, Tersman states that accepting moral principles is a necessary condition for reaching a reflective equilibrium; moral principles provide the evidential structure necessary for maximal coherence.<sup>31</sup> Accordingly, I find it reasonable to assume that Tersman agrees that such principles should by some means satisfy epistemic desiderata in order to provide this evidential structure and avoid constituting mere ad hoc assumptions. Furthermore, in Tersman's refutation of the plurality objection he construes coherence as a flexible notion. Coherence is not a rigid function such that a belief either coheres or does not cohere with some system of beliefs. A belief  $p$  may cohere better or worse with some proposed system of beliefs and how well a belief coheres is as a matter of degree.<sup>32</sup> Thus a belief  $p$  that satisfies the epistemic desiderata to a higher degree than another belief  $q$  could be said to cohere *better* with the system of beliefs in question.

One explanation of coherence that largely corresponds to Tersman's line of reasoning could be the following: a belief  $p$  can gain coherence and thus cohere better with A's system than  $\neg p$ , if  $p$  satisfies the epistemic desiderata to a higher degree than  $\neg p$ . For instance, belief  $p$  may provide more power or add to the systematicity of A's system or perhaps  $p$  is accepted simply to avoid an inconsistency in A's system and so on. Tersman's main assumption is that in a predominant amount of cases (practically all cases),  $p$  and  $\neg p$  cannot both strengthen the coherence of a given system. Either the belief  $p$  strengthens the system and  $\neg p$  does not, or the opposite. In any case we should accept the belief that increases the overall coherence of our system.

### **A Difficulty in Tersman's Account**

The plurality objection fails, according to Tersman, because it is impossible or at least extremely implausible that radically different systems of beliefs can be equally justified. This is so because a) it is unclear how it should be determined that some person in fact has such a system of beliefs, and b) considering such a possibility requires that we in hypothesis ascribe beliefs to that person, which it, on our view, would be irrational to hold. One underlying assumption in this argument is the presupposition that coherence is fixated: if a belief  $p$  coheres to some unspecified degree with a given system of beliefs, then it is not possible for  $\neg p$  to cohere equally well with that given system. However, I think this is a somewhat uncontroversial claim in the sense that any adequate account of coherence would probably include this assumption, but nonetheless it does involve some difficulties.

For instance, and this is my objection to Tersman's account, it seems that in order to determine how well a belief  $p$  in fact coheres with a system of beliefs, it must be determined which epistemic features that must be satisfied – and to what extent.

---

<sup>31</sup> cf. Tersman: Reflective Equilibrium, an Essay in Moral Epistemology - p.50.

<sup>32</sup> cf. Tersman: Reflective Equilibrium, an Essay in Moral Epistemology - p. 104.  
Especially Tersman's assumptions C1. and C2.

Consider the following tentative example: imagine that a person A and a person B assign epistemic first priority to different epistemic desiderata. Say that A regards the *power* of a moral theory as absolutely essential to moral theory-building, while B regards *simplicity* as the most important epistemic feature.

Is it then not possible that A could hold a system of beliefs which would differ if not radically then at least to a very high degree from B's? If the belief *p* coheres very well with A's system of beliefs, does this necessarily decrease the possibility of  $\neg p$  cohering equally well with B's system? I think not: if B assigns epistemic first priority to a different desideratum in his coherent system of beliefs than A, then B may very well accept entirely different beliefs than A – without thereby holding irrational beliefs.<sup>33</sup>

Now, I deliberately wrote a tentative example since I do not intend to suggest that cognitive agents, such as A and B in the above example, should for themselves decide which epistemic features are the most important in moral theory-building. And certainly, coherentists would be right to reject such a claim.

However, I merely intend to demonstrate that unless a rank of priority of the epistemic desiderata is provided, then Tersman's initial assumption cannot be adequately justified. It does seem possible that a belief *p* may cohere with S if we assign first priority to an epistemic desideratum, say *power*, and yet that  $\neg p$  may cohere equally well with S if we assign epistemic first priority to a different epistemic desideratum, say *simplicity*. It seems quite possible that a belief *p* may increase the power of S and nonetheless thereby decrease the simplicity of S.

Coherence as such must be fixated in order to justify the assumption that the better a belief *p* coheres with a system S the worse  $\neg p$  will cohere with S. It must be entirely clear how to determine what a belief's level of coherence is. Without some rank of priority or some guidelines as to what epistemic features we should seek to satisfy, this remains unclear.

Admittedly, my above objection does not pertain directly to the requirement of a meta-justification of the epistemic desiderata.<sup>34</sup> But, in my opinion, the difficulty demonstrated cannot be solved without providing such a meta-justification. This is so because Tersman defends an epistemic interpretation of the method of reflective equilibrium and maintains that a more coherent system of beliefs is better justified than a less coherent system of beliefs.<sup>35</sup> Thus, initially Tersman must in some sense substantiate why the satisfaction of the epistemic desiderata is likely to increase the overall coherence of a system of beliefs. Certainly, if Tersman tried to resolve this difficulty by providing a rank of priority of the epistemic desiderata then it would be quite reasonable to inquire into for instance the justification of the epistemic features and the reasons for settling on exactly those epistemic features etc.

---

<sup>33</sup> If this seems inconceivable, imagine A as holding a moral theory having only one basic principle, say a utilitarian maxim and B as holding a very comprehensive theory of moral rights including a wide range of moral constraints etc. Certainly, the moral beliefs that would be justified in light of these theories are significantly different.

<sup>34</sup> Instead it concerns a rank of priority of the epistemic desiderata. I argued that one such is needed for Tersman's argument to work.

<sup>35</sup> cf. Tersman: Reflective Equilibrium, An Essay in Moral Epistemology – p.11-12.

However, an attentive reader might complain that I have misunderstood the intention of Tersman's argument. Such a reader would perhaps maintain that Tersman's argument merely points out that radically different systems of moral beliefs necessitate radically different systems of non-moral beliefs. That is, the reliability of moral beliefs is intimately related to the reliability of non-moral beliefs.<sup>36</sup> Moreover, it could be argued, Tersman does not try to resolve the problem of diverging moral theories insofar as the non-moral beliefs underlying these theories are largely in agreement. I think this is a reasonable suggestion and I think he is right, if this is in fact the focal point of Tersman's argument. It does seem extremely implausible that two radically different systems of non-moral beliefs could in practice be equally justified. Yet, if this really is the essence of his argument, I also find it rather trivial. Such an argument only rules out extremely different systems of beliefs, and does in no way rule out that very different moral theories could be equally justified. In my opinion, this *is* the very gist of the plurality objection: that very different moral theories can be equally justified, and that this is an unsatisfactory result for a method of moral theorising. Consequently, my objection to Tersman still points out that the plurality objection cannot be entirely dismissed on grounds of holism, radical interpretation, and the principle of charity.

As a result (as in the case of David Brink) in order to make his refutation of the plurality objection work, I think Tersman must also provide a meta-justification of the epistemic desiderata.<sup>37</sup>

### **Sayre-McCord – Coherence not Conducive to Truth**

A quite different approach to the truth and plurality objections is presented by Geoffrey Sayre-McCord. McCord, like Brink and Tersman, defends an epistemic interpretation of the method of reflective equilibrium. However, in contrast to both Brink and Tersman, McCord argues that defending the coherence theory of justification and maintaining moral realism does not commit one to accept that coherence as such is truth-conducive.

In short, McCord's central claim is that to the extent one has evidence for one's beliefs, they will cohere. He does not address the question of meta-justifying epistemic desiderata, but I think he would in fact refuse such a demand. I think so because a meta-justification is required only if one presumes the method of reflective equilibrium to be conducive to truth and McCord explicitly rejects this claim. Consequently, it seems quite reasonable to assume that McCord would view the demand for a meta-justification of the epistemic desiderata as misguided. I think there is a variety of difficulties in McCord's theory, which I will return to later. To begin with, however, I will give a brief presentation of McCord's reasoning.

---

<sup>36</sup> This claim is for instance supported by Brink and is, I think, quite plausible.

<sup>37</sup> However, at this point, the attentive reader might suggest that Tersman should just reject the claim that satisfying epistemic desiderata increases overall coherence. This is obviously an alternative open to Tersman. But, the result of such a rejection is, in my opinion, that a moral theory justified merely by ad hoc principles will cohere just as well as an extremely simple, yet very powerful, and systematic moral theory. Initially, I find this very implausible. Yet the crucial point is that rejecting the epistemic desiderata not only gives rise to equally coherent moral theories – it gives rise to equally coherent, yet *radically different* moral theories. And as a result rejecting epistemic desiderata, in my view, in fact undermines Tersman's own argument.

McCord maintains that for a person A to be justified in holding some belief  $p$ ,  $p$  must cohere with A's additional beliefs. Yet, whether or not  $p$  in fact coheres depends largely on the epistemic perspective of A and the evidence available to A. As an initial premise McCord asserts that certain epistemic requirements must be satisfied in order to yield justification. One of these requirements is what McCord refers to as the *epistemic imperative*.<sup>38</sup> The epistemic imperative demands that for A to be justified in holding a belief  $p$ , the evidence available to A should somehow imply  $p$ . However, the epistemic imperative does not necessitate that A believes everything that the available evidence would allow, but merely that the beliefs that A in fact holds are implied by the evidence available to A.

Furthermore, McCord distinguishes two kinds of justification – *positively* justified beliefs and *permissively* justified beliefs: a belief  $p$  enjoys *positive* justification if the evidence available to A implies  $p$ , while a belief  $p$  can be considered *permissively* justified, if the available evidence does not imply  $\neg p$ . The crucial point of introducing this distinction is the following,

[...] it can explain how the regress might be stopped; it comes to an end if and when we arrive at beliefs that are permissively justified. Second it leaves room for regress-stoppers that, despite their 'regress-stopping' role, might be both over-ridable and underminable; permissively justified beliefs will lose their status when, for instance, new evidence is acquired that tells against them. Third, it avoids saying that among a person's reasons for believing as she does are reasons constituted by considerations that are unavailable to her; whether a belief counts as permissively justified turns only on whether the other things she believes provide, on balance, evidence against the belief. (McCord 1996: p.31)<sup>39</sup>

I will not go into specific details regarding the accurate definitions of permissively justified and positively justified beliefs. I only intend to emphasise that assuming permissively justified beliefs to be effective regress-stoppers makes it quite difficult to contest the plurality objection. If the available evidence does not imply  $\neg p$  we can then consider  $p$  permissively justified. However, if the available evidence does not imply  $\neg p$ , then the evidence cannot imply  $p$  either (since that would imply  $\neg p$ ). So, it seems that depending on the system of beliefs in question both  $p$  and  $\neg p$  can serve as effective regress-stoppers. Obviously, after construing justification in such terms, stating that coherence as such is truth-conducive becomes untenable.

But, on McCord's account coherence is also to be understood in terms of *evidential relations*. For a belief to be justified it must have an evidential relation to one or more additional beliefs in the system. The more evidential relations the particular belief has, the better justified it is. However, it is possible for a belief to cohere with a system of beliefs without having an evidential relation to any other beliefs in the system. Such a belief is permissively justified.

---

<sup>38</sup> cf. McCord: Coherentist Epistemology and Moral Theory - p. 14.

<sup>39</sup> In the first line of the quotation McCord is of course referring to the well known regress argument against coherence theories of justification.

Now, what constitutes an evidential relation is left somewhat unexplained in McCord's theory, which he himself acknowledges.<sup>40</sup> Yet, McCord does state that what counts as an evidential relation can only be established by considering *truth-conducive canons of reasoning*.<sup>41</sup> That is, McCord does not think that an evidential relation between beliefs renders the beliefs more likely to be true. Rather, if the beliefs in a system are thoroughly connected by evidential relations, then the system will cohere. Yet, McCord accepts the possibility of such evidential relations obtaining even though the system of beliefs turn out false,

Obviously, a person might respect the relevant canons of reasoning over time and so hold beliefs that are evidentially related (on this view) and yet, because of lack of evidence, or misleading evidence, actually consistently have evidence, or misleading evidence, actually consistently have evidence for false views. (McCord: 1996 p.38)

The essential point is that there are conditions pertaining to the justification of a system of beliefs that obtain independently of epistemic perspectives and available evidence. These conditions are what McCord refers to as the *evidential consistency, connectedness, and comprehensiveness* of the system.<sup>42</sup> The condition of evidential consistency demands that none of the evidence available to person A tells against A's beliefs. The system's connectedness depends on the degree of positive support (evidential relations) that each belief in the system provides for the additional beliefs: the stronger the overall positive support, the more connected the system will be. Finally, if additional beliefs are added to the system (assuming that the system preserves evidential consistency and connectedness) then comprehensiveness of the system increases – and so does the overall coherence. Yet given factors such as epistemic perspective and available evidence it remains possible to have a thoroughly justified system of beliefs that is in fact false. Whether or not the beliefs a person A holds are in fact justified, turns on conditions that A may be unaware of. And increasing the evidential consistency, the connectedness and the comprehensiveness of one's system of beliefs does not render the system more likely to be true. Yet, if a system of beliefs is true, then these conditions will be satisfied to a very high degree.

### **Sayre-McCord and the Plurality Objection**

McCord does not, as mentioned earlier, intend to contest the truth nor the plurality objections. Rather, he argues that although the coherence of one's system reflects the extent to which one has evidence for one's beliefs, this does not entail that coherence as such is conducive to truth. Claiming that evidential relations will obtain among true beliefs does not commit one to accept that increasing coherence causes one's system of beliefs to be more likely to be true.

This means a coherentist can and should admit that the mere fact that a set of beliefs is coherent provides one with no reason to think they are true, even though, if the beliefs in question are one's own, their relative coherence will

---

<sup>40</sup> 'A full articulation of the coherence theory I've been describing would of course involve developing a theory of what relations count as evidential. And clearly this is not the place to begin that project' (McCord: 1996 – p.43).

<sup>41</sup> cf. McCord: Coherentist Epistemology and Moral Theory – p.38.

<sup>42</sup> Sayre-McCord: Coherentist Epistemology and Moral Theory – p.40.

reflect the extent to which one's evidence gives one reason to think they are true (Sayre-McCord: 1996 – p.47).

The crucial point is that while evidence entails coherence, coherence is not an additional type of evidence.

Initially, I must admit that I find McCord's theory to be very complex and rather unclear. However, whether I lack the sufficient expertise to understand McCord's reasoning or his theory is in fact unclear will remain an open question.

There are, in my view, problems with the method of reflective equilibrium that emerge from McCord's account. Most importantly, seeing as McCord defends an epistemic interpretation of the method of reflective equilibrium, it remains fairly unclear to me what the utility of the method is. Reaching a reflective equilibrium, on McCord's view, provides no reasons for thinking that the moral principles settled upon in the equilibrium are true. The moral principles (and our moral beliefs) are perhaps justified but, as maintained by McCord, this provides no reason to regard those principles as more likely to be true than any other coherent set of moral principles. Seeing as the method of reflective equilibrium may provide a variety of different sets of principles, why at all bother to carry out this process? It seems to me that on McCord's account there is no obvious purpose of utilising the method of reflective equilibrium unless he adopts or combines his theory with e.g. pragmatic considerations. If not, then an explication is needed as to why the method of reflective equilibrium preserves a right as the most satisfying *epistemic* method of moral theorising.

Secondly, I feel somewhat unconvinced that McCord can successfully avoid questions regarding truth. A constitutive factor of justification, as stated by McCord, is the evidential relations connecting beliefs. However, McCord does not provide any clues as to what such an evidential relation amounts to. I find it rather difficult to imagine a satisfactory explication of evidential relations that entirely avoids the notion of truth or, at least, the probability of a belief being true. In other words, if an evidential relation provides justification and this is explained without referring to the probability of some belief being true, then it becomes a fairly challenging task to understand what justification is in McCord's view.

Thirdly, seeing as McCord agrees that the method of reflective equilibrium is designed to uncover justified moral principles, circumventing epistemic desiderata involves difficulties. Certainly, McCord would want the evidential relations connecting beliefs and proposed moral principles to include some epistemic features.<sup>43</sup> If not, then a moral theory comprised of ad hoc principles established by some individual's idiosyncratic moral beliefs would be entirely as justified as any other moral theory. If this in fact is McCord's position, then we are back to my first objection: there is no obvious reason for utilising the method of reflective equilibrium.

However, if McCord concurs that an acceptable evidential relation necessitates the satisfaction of some epistemic features, he then faces the same difficulties as Brink and

---

<sup>43</sup> Here assuming that considered moral beliefs constitute the 'evidence' for proposing some moral principle.

Tersman: neither the truth nor plurality objection can be rejected without a meta-justification of the epistemic desiderata.

In McCord's case one could reasonably ask him the following questions: a) which epistemic desiderata must be satisfied to warrant an acceptable evidential relation? b) why are exactly those desiderata essential in regards to evidential relations and c) to what degree must the epistemic features be satisfied?

## Concluding remarks

In the preceding sections I have attempted to demonstrate that an important aspect of moral theory-building has been somewhat overlooked in the epistemic interpretations of the method of reflective equilibrium. As argued in the above sections, I think that the structure of a moral theory is intimately related to the possible truth of such a theory. In order to ascertain a truth-conducive moral theory by the method of reflective equilibrium, it seems that some standards of satisfactory theory structure are necessary. That is, if a moral theory is to provide plausible moral judgments, the structure of such a theory must also maintain an intuitive plausibility. This claim is (hopefully) supported by my above outline of difficulties in Brink's, Tersman's and McCord's very different suggestions to circumvent the truth and plurality objections.

A moral theorist should hesitate in rejecting considered moral judgments, because such judgments ensure that our theory avoids being a mere stipulation. However, to ascertain a moral theory structure that we intuitively regard as providing adequate justification, some considered moral judgments should probably be rejected. Resolving this tension between moral intuitions and theoretical claims is of course the main purpose of the method of reflective equilibrium; it is certainly the intent that Rawls had in mind. Yet, it must be acknowledged that even if a moral theory were constructed in which all considered moral judgments and moral principles had a perfect fit, this would not necessarily be sufficient. This fit must be guided by some conditions of adequate theory structure and these conditions must in turn be justified. These conditions are of course *the epistemic desiderata*, the epistemic features that we intuitively regard necessary for providing a satisfactory justification.

My fairly discouraging conclusion of this paper is that providing such a meta-justification seems, at least to me, somewhat hopeless. I have absolutely no idea how one would provide an adequate definition of each epistemic desideratum and furthermore explicate the extent to which they must be satisfied. And yet, if a meta-justification of the epistemic desiderata cannot be given, I find that epistemic interpretations of the method of reflective equilibrium still face the plurality and the truth objections: objections that in my opinion are fatal to the method. Perhaps the only reasonable conclusion is that non-epistemic interpretations of the method of reflective equilibrium remain the best available alternative.

Anders Johan Schoubye  
December, 2003

## References & Bibliography

- Audi, Robert. (1998). *Epistemology*. Routledge Contemporary Introductions to Philosophy, London.
- Bonjour, Laurence. (1985). *The Structure and Content of Empirical Knowledge*, Harvard University Press. Cambridge.
- (1999). The Dialectic of Foundationalism and Coherentism. *The Blackwell Guide to Epistemology* p.117-43.
- Brandt, Richard. (1979). *A Theory of the Good and the Right* Clarendon Press. Oxford.
- Brink, David. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge University Press.
- Daniels, Norman. (1996). *Justice and Justification*. Cambridge University Press.
- (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy* – p.256-282
- Davidson, Donald (1984). *Inquiries into Truth and Interpretation*. 2<sup>nd</sup> edition. Clarendon Press, Oxford 2001.
- (1967). Truth and Meaning. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford 2001.
- (1973). Radical Interpretation. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford 2001.
- (1973). In Defence of Convention T. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford 2001.
- (1974). On the Very Idea of a Conceptual Scheme. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford 2001.
- Dworkin, Ronald (1976). The Original Position. *Reading Rawls – Critical Studies of ‘A Theory of Justice’* p. 16-53 (edited by Norman Daniels).
- Everitt, Nicholas & Fisher, Alec. (1995). *Modern Epistemology – A New Introduction*. McGraw-Hill.
- Freeman, Samuel (ed.). (2003). *The Cambridge Companion to Rawls*. Cambridge University Press.
- Greco, John (ed.) & Sosa, Ernest (ed.). (1999). *The Blackwell Guide to Epistemology*. Blackwell. Malden, Massachusetts.
- Hare, R.M. (1973). Rawls’ Theory of Justice. *Reading Rawls – Critical Studies of ‘A Theory of Justice’* p.81-108.
- Harnish, Robert M. (1994). *Basic Topics in the Philosophy of Language*. Harvester Wheatsheaf.
- Kagan, Shelly. (1989). *The Limits of Morality*. Clarendon Press, Oxford.
- Kappel, Klemens. (2003). ‘Is there a Meta-Justification of Reflective Equilibrium?’ Unpublished draft.
- Kirkham, Richard. (1995). *Theories of Truth – A Critical Introduction*. MIT Press, Cambridge Mass.
- Lyons, David (1973). The Nature and Soundness of the Contract and Coherence Arguments. *Reading Rawls – Critical Studies of ‘A Theory of Justice’* p. 141-169.
- Rawls, John. (1971). *A Theory of Justice*. Oxford University Press
- Quine, W.V.O. (1960). *Word and Object*. MIT Press, Cambridge Mass.
- Sayre-McCord, Geoffrey. (1996). Coherentist Epistemology and Moral Theory. *Moral Knowledge?: New Readings in Moral Epistemology*
- Scanlon, T.M. (2003) Rawls on Justification. *The Cambridge Companion to Rawls*
- Tersman, Folke. (1993). *Reflective Equilibrium – An Essay in Moral Epistemology*. Almqvist & Wiksell International, Karlshamn
- Timmons, Mark (ed.) & Sinnott-Armstrong, Walter (ed.). (1996). *Moral Knowledge?: New Readings in Moral Epistemology*. Oxford University Press.
- Williams, Michael. (1991) *Unnatural Doubts*. Blackwell Publishers Inc, Cambridge

Sign count: 69257

Pages: 28.9