

EPISTEMOLOGIA CONTEMPORÂNEA EM DEBATE

# Self-Knowledge and Epistemic Virtues: Between Reliabilism and Responsibilism

*Autoconhecimento e virtudes epistêmicas: entre o  
confiabilismo e o responsabilismo*

\*César Schirmer dos Santos

---

**Abstract:** This paper is about the role of self-knowledge in the cognitive life of a virtuous knower. The main idea is that it is hard to know ourselves because introspection is an unreliable epistemic source, and reason can be a source of insidious forms of self-deception. Nevertheless, our epistemic situation is such that an epistemically responsible agent must be constantly looking for a better understanding of her own character traits and beliefs, under the risk of jeopardizing her own status as a knower, ruining her own intellectual life.

**Keywords:** Self-Knowledge. Virtue Epistemology. Intellectual Virtues. Introspection. Reason.

**Resumo:** O tema deste artigo é o papel do autoconhecimento na vida cognitiva de um sujeito que mereça ser considerado possuidor de conhecimento. A ideia central é que é difícil alcançar o conhecimento de si, pois a introspecção é uma fonte de crenças pouco confiável, e a razão pode ser a fonte de formas extremas de autoengano. Apesar disso, a nossa situação é tal que um agente epistêmico responsável deve buscar constantemente uma melhor compreensão dos seus próprios traços de caráter e crenças, sob o risco de pôr em perigo o seu status de sujeito possuidor de conhecimento.

**Palavras-chave:** Autoconhecimento. Epistemologia das virtudes. Virtudes intelectuais. Introspecção, Razão.

---

\* PhD. in Philosophy, Associate Professor at the Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. Post address: Av. Roraima, 1000 – Prédio 74A, Sala 305; Departamento de Filosofia – UFSM; Cidade Universitária – Bairro Camobi; CEP: 97.105-900, Santa Maria, RS, Brazil. <[cesar.santos@ufsm.br](mailto:cesar.santos@ufsm.br)>.



## Introduction

In this paper, I use virtue epistemology for the inquiry of self-knowledge<sup>1</sup>. I will argue that self-knowledge never can be fully obtained, but must be always looked for, because an epistemic agent has to try hard to know herself, since that's the price of intellectual virtue. I will proceed as follows. First, I will present the basics of virtue epistemology, and the main theories of self-knowledge. Then I will use virtue epistemology for the analysis of self-knowledge.

### 1 Virtue Epistemology

Virtue epistemology is an approach to traditional epistemological problems proposed by Ernest Sosa in his seminal paper "The Raft and the Pyramid"<sup>2</sup>. The idea was inspired by the Aristotelian approach to ethics on the rise since the 1950s<sup>3</sup>. In virtue ethics, a laudable action is an action produced through a process properly related to a *moral* virtue. By analogy, in epistemology, a praiseworthy belief is a belief produced by a process involving an *intellectual* virtue.

Having in mind specifically Sosa's seminal paper, and the philosophical debates of its time, virtue epistemology was presented, some twenty-five years ago, as an alternative both to epistemological foundationalism and epistemological coherentism. Sosa describes the main dilemma of the epistemologists of these days as the choice "between the solid security of the ancient foundationalist pyramid and the risky adventure of the new coherentist raft"<sup>4</sup>. As a way out, Sosa proposes that:

---

<sup>1</sup> Traditionally, the question of self-knowledge is reduced to the question of the knowledge of one's own occurrent mental states. Examples of traditional questions about self-knowledge are 'How do I know that I am in pain?', and 'How do I know that I'm thinking that p?'. More consequential and interesting examples would be, for instance, 'Am I a good teacher?', or 'Am I a racist?'; see CASSAM, Quassim, *Self-Knowledge for Humans*, Oxford: Oxford University Press, 2014. As I see it, responsibility can be used to explain traditional cases of self-knowledge and to explore the more interesting cases of self-knowledge.

<sup>2</sup> See SOSA, Ernest, The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge, in: *Midwest Studies in Philosophy*, 5 (1980), p. 3-25.

<sup>3</sup> See KVANVIG, Jonathan L., Virtue Epistemology, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), *The Routledge Companion to Epistemology*, London: Routledge, 2011, p. 199-207.

<sup>4</sup> See SOSA, Ernest, The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge, op. cit., p. 3. The idea behind Sosa's metaphors is that, on the one hand, while foundationalism is the search for foundational beliefs, promising solid grounds (but opening the possibility of infinite regress), virtue epistemology proposes foundational personal virtues (the character traits shaping the practices of the agent), instead of beliefs, as the ground for knowledge. On the other hand, while coherentism is blind or at least short-sighted for the relation between truth and belief, a virtue epistemologist sees as virtuous the personal skills or character traits conducive to truth. So, virtue epistemology is a hope in the search for solution of old difficulties in the field of epistemology.

[...] primary justification would apply to *intellectual* virtues, to stable dispositions for belief acquisition, through their greater contribution toward getting us to the truth. Secondary justification would then attach to particular beliefs in virtue of their source in intellectual virtues or other such justified dispositions<sup>5</sup>.

So, key to Sosa's proposal is to describe the epistemological concept of justification as a concept that applies, first of all, to some personal capacities or dispositions of believers (or knowers). Sosa dubs these capacities as *intellectual virtues*. The main demand of the original proposal, still present in contemporary virtue theories of knowledge, is to apply the concept of justification primarily to the cognitive faculties or character traits of believers, secondarily to their beliefs<sup>6</sup>. 'Justified' is, primarily, the process of belief production, secondarily, the output or product of the process, which is the belief. This shift in the focus of the concept of justification, from beliefs to believers, is very important for the works of all virtue epistemologists, including the early reception of Sosa's work by Lorraine Code (1984)<sup>7</sup>. But, as we will see, while Sosa's proposal involves reliable dispositions<sup>8</sup>, central to Code's proposal is the concept of epistemic responsibility, since virtue epistemology comes in two versions, virtue reliabilism and virtue responsibilism<sup>9</sup>.

## 2 Two Kinds of Virtue Epistemology

As we saw, virtue epistemology is proposed as an analog, in the field of epistemology, for virtue ethics in the field of practical philosophy. But how are we supposed to understand the analogy? In the field of moral philosophy, a virtue ethicist says that a virtuous or excellent action is the output of a practical process conducted by a generous, courageous, honest, temperate (and so on) character. In the appropriations of the epistemologists, the analogy has two readings:

<sup>5</sup> See SOSA, Ernest, *The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge*, op. cit., p. 23, author's italics.

<sup>6</sup> Goldman identifies "the concept of justified belief with the concept of belief obtained through the exercise of intellectual virtues (excellences)"; see GOLDMAN, Alvin, *Epistemic Folkways and Scientific Epistemology*, in: *Philosophical Issues*, 3 (1993), p. 274 (271-285).

<sup>7</sup> See CODE, Lorraine, *Toward a 'Responsibilist' Epistemology*, in: *Philosophy and Phenomenological Research*, 45:1 (1984), p. 29-50.

<sup>8</sup> See SOSA, Ernest, *The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge*, op. cit., p. 23.

<sup>9</sup> See BAEHR, Jason, *Character, Reliability and Virtue Epistemology*, op. cit., p. 193: "Standard characterizations of virtue epistemology divide the field into two camps: virtue reliabilism and virtue responsibilism. Virtue reliabilists think of intellectual virtues as reliable cognitive faculties or abilities, while virtue responsibilists conceive of them as good intellectual character traits."

- (a) *Reliabilism*: a virtuous (or excellent) belief is the product of an epistemic process that reliably conduces to truth, such as perception, memory, inference and introspection. (I will speak more about the reliability of introspection below.) The most prominent virtue reliabilists are Ernest Sosa (1980)<sup>10</sup>, Alvin Goldman (1993)<sup>11</sup> and John Greco (2010)<sup>12</sup>.
- (b) *Responsibilism*: a virtuous (or excellent) belief is the product of an epistemic process conducted by a fair-minded, open-minded, intellectually careful, intellectually courageous (and so on) agent. The most prominent virtue responsibilists are Lorraine Code (1984)<sup>13</sup>, James Montmarquet (1993)<sup>14</sup> and Linda Zagzebski (1996, 2000)<sup>15</sup>.

In summary, on the one hand, virtue reliabilists assess as virtuous (or excellent) the subpersonal cognitive skills (or faculties) of the agent. On the other hand, virtue responsibilists assess as epistemically virtuous personal-level character traits such as intellectual responsibility, fair-mindedness, open-mindedness, intellectual carefulness, intellectual courage, intellectual humility, intelligence, empathy, integrity, perseverance, faith in reason, and so on<sup>16</sup>.

## 2.1 Reliabilism

Sosa's approach to virtue epistemology is reliabilist. Roughly, reliabilism in epistemology is the view that knowledge is true belief "obtained by a reliable process"<sup>17</sup>. Since the 1970s, many reliabilist

<sup>10</sup> See SOSA, Ernest, *The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge*, op. cit., p. 3-25.

<sup>11</sup> GOLDMAN, Alvin, *Epistemic Folkways and Scientific Epistemology*, op. cit., 1993.

<sup>12</sup> See GRECO, John, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge, UK: Cambridge University Press, 2010. Sometimes, to classify virtue epistemologists is not so easy. On the list of virtues, GOLDMAN, Alvin, *Epistemic Folkways and Scientific Epistemology*, op. cit., p. 275, says that "the virtues include belief formation based on sight, hearing, memory, reasoning in certain 'approved' ways, and so forth." As I see it, that's a mixed list, reliabilist and responsibilist, since perception and memory are subpersonal faculties, but reasoning in 'approved' ways is a personal-level output.

<sup>13</sup> See CODE, Lorraine, *Toward a 'Responsibilist' Epistemology*, op. cit., p. 29-50.

<sup>14</sup> See MONTMARQUET, James, *Epistemic Virtue and Doxastic Responsibility*, Lanham, MD: Rowman & Littlefield, 1993.

<sup>15</sup> See ZAGZEBSKI, Linda, *Virtues of the Mind*, Cambridge, UK: Cambridge University Press, 1996; EADEM, *From Reliabilism to Virtue Epistemology*, in: AXTELL, Guy (ed.), *Knowledge, Belief and Character: Readings in Virtue Epistemology*, Lanham: Rowman & Littlefield, 2000.

<sup>16</sup> See CODE, Lorraine, *Toward a 'Responsibilist' Epistemology*, op. cit., p. 29-50; AXTELL, Guy, Introduction, in: AXTELL, Guy (ed.), *Knowledge, Belief, and Character: Readings in Virtue Epistemology*, Lanham: Rowman & Littlefield, 2000; BAEHR, Jason, *Character, Reliability and Virtue Epistemology*, op. cit., p. 193.

<sup>17</sup> See RAMSEY, Frank, *Knowledge*, in: RAMSEY, Frank, *The Foundations of Mathematics and Other Logical Essays*, London: Routledge and Kegan Paul, 1931, p. 110, *apud* COMESAÑA, Juan, Reliabilism, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), *The Routledge Companion to Epistemology*, London: Routledge, 2011, p. 176.

theories were developed to solve Gettier (1963)<sup>18</sup> puzzles about the analysis of knowledge as justified true belief<sup>19</sup>. The usual reliabilist strategy is to explain the concept of justification *a la* Ramsey: a true belief is justified when it's the output of a reliable process<sup>20</sup>. Bringing the two ends (intellectual virtues and reliabilism) together, Sosa's theory of knowledge says that knowledge is the output of a process that is justified because it involves intellectual virtues, and intellectual virtues are reliable. So, generalizing Sosa's view, a virtue reliabilist says that knowledge is achievement from intellectual virtues, or excellences<sup>21</sup>.

## 2.2 Responsibilism as a Criticism of Reliabilism

I will first present a simpler form of responsibilism. Below I will fill some blanks. For a responsibilist, "intellectual virtues are personality traits that a person who desires the truth would want to have"<sup>22</sup>. It means that an intellectual virtue is a personal-level disposition that is truth-conducive.

Lorraine Code is one of the main epistemologists in the responsibilist side. Code's project is to implement Sosa's view, making explicit the importance of the knower in the "account of the nature and conditions of knowledge"<sup>23</sup>. There is common ground between Sosa's and Code's theories, since in both theories the goal is to apply the concept of justification primarily to believers (or knowers), secondarily to beliefs (or knowledge). But their paths come apart on the issue of the proper characterization of an intellectual virtue. Code argues that since reliability without responsibility is not enough to generate a virtuous epistemic relation to the world<sup>24</sup>, the intellectual virtues have to be located at the personal level:

<sup>18</sup> See GETTIER, Edmund, Is Justified True Belief Knowledge?, in: *Analysis*, 23 (1963), p. 121-123.

<sup>19</sup> See COMESAÑA, Juan, Reliabilism, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), op. cit., p. 176.

<sup>20</sup> Id. *ibid.*, p. 177.

<sup>21</sup> See GRECO, John, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, 2010.

<sup>22</sup> See GRECO, John and TURRI, John, Virtue Epistemology, in: ZALTA, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), URL = <<http://plato.stanford.edu/archives/win2013/entries/epistemology-virtue/>>, accessed Nov. 14, 2014.

<sup>23</sup> See CODE, Lorraine, Toward a 'Responsibilist' Epistemology, op. cit., p. 30.

<sup>24</sup> More recently, PRITCHARD, Duncan and TURRI, John, The Value of Knowledge, in: ZALTA, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), URL = <<http://plato.stanford.edu/archives/spr2014/entries/knowledge-value/>>, accessed Nov. 11, 2014, objected to virtue reliabilism that sometimes reliability is not enough to knowledge, sometimes it's not necessary. I cannot deal with this question in this paper. See KELP, Christoph, In Defense of Virtue Epistemology, in: *Synthese*, 179 (2011), p. 409-433.

To be intellectually virtuous is not just to have a good score in terms of cognitive endeavours that come out right. It is much more a matter of orientation toward the world, and toward one's knowledge-seeking self, and other such selves as part of the world. Virtues, both moral and intellectual, have much more to do with the manner of relation to the world than with the 'content' of particular actions or knowledge claims<sup>25</sup>.

Code's critique is devastating for any reliabilist project in *virtue* epistemology that eschews the concept of responsibility. Reliable is the process which output has a high truth-ratio<sup>26</sup>. But, Code says, "a good score" is not enough for intellectual virtue. That's a *reductio* of virtue reliabilism without epistemic responsibility. Parsing the argument:

- (1) If virtue reliabilism is right, then production of a high ratio of true beliefs is enough to intellectual virtue.
- (2) But production of a high ratio of true beliefs without epistemic responsibility is not enough to intellectual virtue.
- (3) Therefore, virtue reliabilism is not right.

Since Code's fundamental intellectual virtue is epistemic responsibility<sup>27</sup>, what is lacking in virtue reliabilism is the responsibility of the agent. When the agent is epistemically responsible, she is intellectually virtuous. Thus, Code's formula for virtue epistemology is:

S knows that p iff the true belief that p is actively obtained by an epistemically responsible (intellectually virtuous) process.

In sum, according to Code's proposal, an intellectual virtue is a full-blood character trait (instead of a merely subpersonal cognitive faculty) that contributes for the justification or warrant of a belief<sup>28</sup>. Code's theory amounts to attribute justification to epistemic processes conducted with epistemic responsibility. Having in mind two things, first, that according to virtue epistemology an agent S achieves knowledge when her true belief is the product of an excellent (or virtuous) cognitive process, second, that epistemic responsibility is the fundamental intellectual virtue, S knows when her true belief is the output of her epistemic responsibility<sup>29</sup>.

<sup>25</sup> See CODE, Lorraine, Toward a 'Responsibilist' Epistemology, op. cit., p. 41.

<sup>26</sup> See COMESAÑA, Juan, Reliabilism, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), op. cit., p. 182.

<sup>27</sup> See CODE, Lorraine, Toward a 'Responsibilist' Epistemology, op. cit., p. 42.

<sup>28</sup> See BAEHR, Jason, Character, Reliability and Virtue Epistemology, op. cit., p. 194.

<sup>29</sup> This point is related to the question of the *value* of knowledge, but I will not discuss the issue here. See GRECO, John, The Value Problem, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), *The Routledge Companion to Epistemology*, London: Routledge, 2011, p. 219-231.

### 2.3 Responsibilism as a Kind of Reliabilism

Code's criticism of reliabilism presupposes the incompatibility between reliability and responsibility. I guess that her motivation is the view of reliability as a too mechanistic concept: "We would speak of a "reliable" computer, but not of a "responsible" one"<sup>30</sup>. Linda Zagzebski<sup>31</sup>, by her turn, depicts responsibility as a kind of reliability. Zagzebski's view can be parsed like that:

- (1) S is motivated to get to the truth.
- (2) S acts in reliable ways because of that motive.
- (3) S is successful because of the motive and the reliable processes occasioned by the motivation.
- (4) Then S's epistemic state worth more than true belief – it's knowledge.

I think that Zagzebski's view is correct. As a matter of fact, we usually classify responsible people as reliable people. Computers can be just reliable, but people can be reliable and have to be responsible. Moreover, Code's criticism of virtue reliabilism is that reliability is no condition for knowledge – it's not necessary, and it's not sufficient. But how could the *moral* concept of responsibility generate *epistemic* outputs by itself? It can't. Responsible cognitive agents produce knowledge because they are reliable.

In summary, a virtue epistemologist says that a belief is justified if it is the output of a reliable process. A reliable process involves an intellectual virtue as a stable mean for the production of true beliefs. Some virtue epistemologists, the virtue reliabilists, say that cognitive faculties such as perception, memory, reasoning and introspection are intellectual virtues. Some other virtue epistemologists, the virtue responsibilists, disagree. They say that the real intellectual virtues are the character traits conducive to truth, such as impartiality, awareness of one's own fallibility, perseverance in the investigation of unpopular ideas, and so on. Nevertheless, the two varieties of virtue epistemology are not incompatible. It's possible to put the two together in a single theory, and it's good to do that, because the resulting theory is stronger. In this resulting mixed theory, justified is, primarily, any truth-conducive cognitive faculty or character trait. If they are truth-conducive, then they are reliable. I will use this mixed theory for the analysis of self-knowledge.

So much for tool-sharpening.

<sup>30</sup> See CODE, Lorraine, *Toward a 'Responsibilist' Epistemology*, op. cit., p. 40.

<sup>31</sup> See ZAGZEBSKI, Linda, *From Reliabilism to Virtue Epistemology*, in: AXTELL, Guy (ed.), op. cit., 2000.

### 3 Self-Knowledge

There are two main groups of accounts for the knowledge of one's own mental states, namely empiricism and rationalism<sup>32</sup>. The main empiricist theory is Armstrong's inner sense theory<sup>33</sup>. The main rationalist theory is Moran's transparency theory<sup>34</sup>.

#### 3.1 Introspection

Introspection or inner sense is the ability to quasi-perceptually access one's own mental states<sup>35</sup>, which means that introspection is the internal analog for perceptual knowledge<sup>36</sup>. We perceive our own mental states by the inner sense, just as we perceive the external world by the traditional five senses.

The faculty of introspection figures in virtue reliabilists lists of intellectual virtues, which implies that introspection is seen as a reliable source of true beliefs. Nevertheless, as we will see, the question of the reliability of the 'inner sense' is at least polemical, if not plainly hopeless.

In inner sense theory, the analogy with the external senses explains why introspection is reliable, but not infallible<sup>37</sup>. Under some circumstances, inner sense can give false beliefs about our own mental states in the same way that vision can give false beliefs about distant round (or square)

<sup>32</sup> See GERTLER, Brie, *Self-knowledge*, London: Routledge, 2011.

<sup>33</sup> See ARMSTRONG, D. M., Is Introspective Knowledge Incorrigible?, in: *The philosophical Review*, 72:4 (1963), p. 417-432.

<sup>34</sup> See MORAN, Richard, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton: Princeton University Press, 2001.

<sup>35</sup> Sometimes, 'introspection' is viewed as a synonym for 'self-knowledge.' But this kind of view just begs the question. A fuller exposition of problems related to the notion of introspection would require the proper distinction between at least three different varieties of phenomena: introspection as one single thought event, introspection as meta-cognition and introspection as narrative about oneself. Unfortunately, this kind of investigation is beyond the limits of this paper.

<sup>36</sup> An empiricist such as ARMSTRONG, D. M., Is Introspective Knowledge Incorrigible?, op. cit., (1963), p. 417-432, defends that knowledge of our own mental states is similar to knowledge of the external world, since all kinds of knowledge are achieved by the senses: mental states by inner sense, external world by external senses (see GERTLER, Brie, *Self-knowledge*, 2011). Inner sense theory has its own epistemological problems. E.g., our external senses are enabled by external sensory organs. By analogy, do we have (at least) one internal sensory organ in our brains? I will not address this question in this paper.

<sup>37</sup> That's the difference between Armstrong's theory of introspection and the theory of acquaintance of RUSSELL, Bertrand, *The Problems of Philosophy*, Oxford: Oxford University Press, 1912, p. 76. Introspection is fallible, Russell's acquaintance is infallible. As a subtopic (a new layer of confusion) related to the word 'acquaintance,' but unrelated to infallibility, DRETSKE, Fred, Skeptical Doubts About Self-Knowledge, in: BERNECKER, Sven and PRITCHARD, Duncan (eds.), *The Routledge Companion to Epistemology*, London: Routledge, 2011, p. 425-432, dubs as 'acquaintance' the usual (and fallible) familiarity that an agent has with her own mental states.



towers<sup>38</sup>. But inner sense also can be truth-conducive in the same way that vision can be truth-conducive<sup>39</sup>. The problem is that, empirically, the truth-ratio for beliefs generated by inner sense is really bad, which means that introspection is not reliable. So, it's not an intellectual virtue. Having that information in mind, can introspection be the source of self-knowledge? As a matter of *a priori* metaphysical possibility, yes it can. But, as a matter of historical (biological, developmental) contingency, we have lots of reasons to say that no, it can't. The problem is that the nice and reliable introspection of the philosophers can't be found by empirical research. The following *a priori* conditions are not *empirically* satisfied by some introspective faculty of real people.

Introspective beliefs are epistemically justified only if they are both safe and virtuous: safe in that the belief would then be held only if true; virtuous in that it derives from the exercise of an intellectual virtue, an internal condition, as explained, that enables the subject to attain certain desirable intellectual ends, pre-eminently that of truth<sup>40</sup>.

Epistemic safety is not my concern in this paper, since my focus is only on epistemic virtue. According to Sosa, an introspective belief is a belief which content is generated by the faculty of introspection. This belief is virtuous if the faculty of introspection enables the agent to attain an epistemic goal such as truth or understanding.

Is it metaphysically possible that an agent (extraterrestrial, perhaps) obtain a true belief about her own mental states through the virtuous faculty of introspection? Yes, it is, since there is no contradiction involved in the proposal. But is it empirically probable that a *human* agent obtain this kind of knowledge by this kind of source? Hardly. Rarely.

The empirical findings of Nisbett and Wilson<sup>41</sup> about the unreliability of the faculty of introspection are bad news for Armstrong and Sosa. There

<sup>38</sup> In contemporary theory, in the most usual construction, inner sense is the faculty that generates second-order beliefs about first-order mental states. That's Higher-Order Theory (HOT). Some other philosophers construct inner sense as Higher-Order Perception (HOP). And there are still other philosophers that construct inner sense as Self-Representational (SR) belief or perception (see KIDD, Chad, The Idols of Inner-Sense, in: *Philosophical Studies* (online), October 28, 2014. DOI: [10.1007/s11098-014-0389-6](https://doi.org/10.1007/s11098-014-0389-6)). In the history of philosophy, Descartes saw inner sense as a first-order perception of 'internal' states of the body, such as hunger and thirst (see BAKER, Gordon and MORRIS, Katherine J., *Descartes' Dualism*, London: Routledge, 1996), which means that Descartes was an empiricist about inner sense.

<sup>39</sup> The theory is naturalistic, since it says that a subpersonal cognitive faculty explains first-person authority. Psychologists and neuroscientists do research it empirically.

<sup>40</sup> See SOSA, Ernest, *Consciousness and Self-Knowledge*, in: GERTLER, Brie (ed.), *Privileged Access: Philosophical Accounts of Self-Knowledge*, Farnham: Ashgate, p. 261.

<sup>41</sup> See NISBETT, Richard E. and WILSON, Timothy DeCamp, Telling More Than We Can Know: Verbal Reports on Mental Processes, in: *Psychological Review*, 84:3 (1977), p. 231-259.

is empirical evidence that we have little or no access to a higher order cognitive process such as introspection<sup>42</sup>. When people are prompted to report their own states of mind, “their reports are based on a *priori*, implicit causal theories, or judgements about the extent to which a particular stimulus is a plausible cause of a given response”<sup>43</sup>. So we don't introspect ourselves. Instead, we create theories for the contents of our own minds that are plausible because they cohere with our behavior and the contextually observed situation of the external world. As a matter of fact, these *ex post facto* theories are reliable<sup>44</sup>. But they are not introspected<sup>45</sup>.

Let's take stoke. As a matter of contingent fact about the human condition, we can't know ourselves by introspection. Thus, for we human beings, introspection isn't a reliable ground for belief. Introspection, for we human beings, isn't an intellectual virtue. So we have to be skeptical about the possibility of human self-knowledge grounded on the outputs of introspection.

## 3.2 Rationalism

We can't know ourselves by self-observation. Can we know ourselves by the sheer endorsement of our own reasons? Let's see Moran's theory about the normal mode of belief-formation<sup>46</sup>. It's based on a famous passage from the work of Gareth Evans:

[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. 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?'<sup>47</sup>.

Transparency is the feature of the first-person perspective that explains why, when S self-ascribes a belief, S looks towards the external

---

<sup>42</sup> Id. *ibid.*, p. 231. A cognitive process is a process “mediating the effects of a stimulus on a response”.

<sup>43</sup> *Ibid.*

<sup>44</sup> *Ibid.*

<sup>45</sup> Example: It's Saturday morning, and eight years old Sam just want to watch cartoons on TV. But there is no cartoons because every TV channel is broadcasting the funeral of a princess. Sam is mad and upset. Her mother asks why she is so upset ('Oh darling, it's because your beloved princess died?'), Sam answers that's because there is no cartoons to watch. Her answer is a plausible and reliable *ex post facto* theory.

<sup>46</sup> Moran's theory of self-knowledge involves an argument against the self-observational model. Nevertheless, for the sake of brevity, I will skip it, since we already have sufficient grounds for the rejection of inner sense theory.

<sup>47</sup> See EVANS, Gareth, *The Varieties of Reference*, Oxford: Oxford University Press, 1982, p. 225.

world, not inwards herself<sup>48</sup>. According to Moran's model, one knows one's own attitudes by knowing one's own reasons, with no need of self-observation<sup>49</sup>. Since one's own propositional attitudes have to be tied to one's own reasons, S can know S's own mind just assessing S's own reasons<sup>50</sup>. If asked about the weather, S should ground her answer in (something like) S's visual experience of rain outdoors, and report it's raining. This report would have two epistemic roles. First, it would report what S knows about the weather. Second, it would report what S knows about S's own mind: that S believes it's raining. S doesn't need to introspect to know it about herself, because the consideration of S's own reasons is enough, and argument is required to collapse a reasoning process into an introspective process. So, according to Moran's theory, self-knowledge is special for a normative reason: the rights and responsibilities of a rational agent are enough for the agent's knowledge of her own mental states. This model is different from the observational model, in which self-knowledge is special by the way of knowledge involved: self-observation. The transparency method is intrinsically first-personal, since it works just for S's knowledge of S's own propositional attitudes, not working, for instance, for my knowledge of your mental states or your knowledge of my own mental states.

In sum, Moran says that S's endorsing of p (the action of the endorsement) is enough for S's knowledge of S's own belief that p. So, in Moran's theory, transparency is a *method* for the achievement of self-knowledge. The problem is that belief and endorsement are very different phenomena<sup>51</sup>. Imagine a case involving Müller-Lyer illusion. Moved by reasons, S could wholly endorse the proposition that the two lines have

<sup>48</sup> In Moran's theory, the transparency method is required for self-knowledge because it's constitutive of the concept of a person to see herself as a center of causality moved by reasons (see SCHMID, Ulla, *Where Individuals Meet Society: The Collective Dimensions of Self-Evaluation and Self-Knowledge*, in: ZIV, Anita K.; LEHRER, Keith; SCHMID, Hans B. (eds.), *Self-Evaluation: Affective and Social Grounds of Intentionality*, Dordrecht: Springer, 2011, p. 266). Something capable of detecting a flow of mental states about the world without taking a stance would not be an agent.

<sup>49</sup> GERTLER, Brie, op. cit., 2011, objects that the transparency theory explains the formation of new intentional attitudes, saying nothing about the attitudes the subject already endorses. But this objection begs the question, since it put apart from the realm of knowledge the attitudes achieved by transparency (see SCHMID, Ulla, *Where Individuals Meet Society: The Collective Dimensions of Self-Evaluation and Self-Knowledge*, in: ZIV, Anita K.; LEHRER, Keith; SCHMID, Hans B. (eds.), op. cit., p. 263).

<sup>50</sup> See SILVA FILHO, Waldomiro and DAZZANI, Maria Virginia, *Self, Sign, Belief, and Agency*, in: *Interações*, 24 (2013), p. 84: "(Condition of Transparency:) The question of the first person present tense about his/her own belief ("Do I believe that p?") is answered in relation to the same reasons that justify an answer to the corresponding question about the world (about the truth of p)."

<sup>51</sup> See MCGEER, Victoria, *The Moral Development of First-Person Authority*, in: *European Journal of Philosophy*, 16:1 (2007), p. 81-108.

the same length, without fully believing it. Belief is much more complex than endorsement.

### 3.3 Partial Conclusion About Empiricism and Rationalism Concerning Self-Knowledge

Empirical research on introspection show it as an utterly unreliable faculty. There is no minimal guarantee that an agent S relying on introspection would obtain genuine knowledge, because we human beings are prone to call 'introspection' our own *ex post facto* rationalizations. Introspection is no intellectual virtue.

Moran transparency theory is restricted only to beliefs. It gives no account for other propositional attitudes as desire or fear. It can't explain the knowledge of one's own feelings and emotions. Moran's account of belief doesn't work properly because belief is unduly reduced to endorsement.

## 4 Back to Responsibilism

Introspection can't be an epistemic ground for self-knowledge because it's not an intellectual virtue, and Moran's rationalism failed for its own problems<sup>52</sup>. The solution can be away, at the core of virtue responsibilism. Crucial to virtue epistemology is to ground justification primarily on the believer (or knower). But, in the case of self-knowledge, we can't ground it on introspection or other intellectual virtues of virtue reliabilism. Can we ground it on the intellectual virtues of virtue responsibilism?

Code's central intellectual virtue is epistemic responsibility. S knows that p iff the true belief that p is actively obtained by an epistemically responsible process. The knower has to choose, "and is accountable for these choices"<sup>53</sup>. Being epistemically responsible, S substantiates her beliefs "with intellectual honesty and due care", "knowing what cognitive ends are worth pursuing"<sup>54</sup>. Intelligence also is an intellectual virtue. Being intelligent, S "attempts to look at the situations clearly, carefully, so as not to be *unduly* swayed by affectivity"<sup>55</sup>. Prudence is a third intellectual (and moral) virtue. Being prudent, S has "a sense of one's limitations"<sup>56</sup>.

---

<sup>52</sup> Perhaps reason can do the work in some kind of inferentialism about self-knowledge (see CASSAM, Quassim, *Self-Knowledge for Humans*, 2014), but that is not the subject of this paper.

<sup>53</sup> See See CODE, Lorraine, Toward a 'Responsibilist' Epistemology, op. cit., p. 40.

<sup>54</sup> Id. *ibid.*, p. 41.

<sup>55</sup> *Ibid.*, p. 42; author's italics.

<sup>56</sup> *Ibid.*, p. 43.

In summary, an intellectually virtuous believer (or knower) actively looks for meaningful grounds for her beliefs, being honest and careful. She is aware of the perils of bias caused by emotional attachments, but she also stays alert for the vice of lack of proper empathy for the others. Lastly, she knows that she can't know everything, and for this reason she puts her intellectual energy in questions of collective importance and projects that can be finished or at least carried out with her own intellectual forces. This portrait of an intellectually virtuous agent is the portrait of a virtuous seeker of knowledge and self-knowledge:

Intellectual virtue is a matter, too, of orientation toward one's knowledge-seeking self. What I mean by this is that, for the intellectually virtuous, self-knowledge is as important as, and indeed complementary to, knowledge of the world. [...] I think it is just because self tends to obtrude so insistently in all human activity, all attempts to be 'objective,' that self-knowledge is essential. It is important to know [...] oneself if one is to achieve a just estimation of the degree to which one does know, believe justifiably, deceive oneself, or fail in epistemic responsibility. [...] Here [...] because of the impossibility of separating knower and known, the difficulties of assessment are great. Nonetheless, despite the apparent circularity in such undertaking, one must strive to know oneself if one is responsibly to assess the degree of one's own responsibility, both epistemic and moral. One cannot improve upon it if one does not know it. Reflexivity is both a fact of human capacity and (potentially, at least) a value<sup>57</sup>.

In Code's view, to know the world the agent has to be intellectually virtuous. But, to be intellectually virtuous, the agent must at least try hard to know herself. Self-knowledge is not a special kind of knowledge, it is one of the epistemic goals for any knower, because to neglect self-knowledge is to lack intellectual virtue.

So responsibilism holds a place for self-knowledge. One must try to know one's own beliefs and character traits to become more intellectually virtuous. In a Socratic note, we could say that striving is enough, achieving is not required. Self-knowledge is a task to be performed, not a fact to be explained. Self-knowledge is a practical possibility to be sought by the virtuous<sup>58</sup>.

<sup>57</sup> Ibid.

<sup>58</sup> I would like to thank to the audiences of 1º Colóquio Internacional de Epistemologia Analítica (September 2-5, 2014, Santa Maria, RS, Brazil), IV International Meeting on Skepticism: The Value of Understanding (November 17-19, 2014, Salvador, BA, Brazil), V Jornada Nacional de Pesquisa na Pós-Graduação em Filosofia da UFSM (November 26-28, 2014, Santa Maria, RS, Brazil). Special thanks to Hilan Bensusan, Eros M. de Carvalho, Juan Comesaña, Rogerio S. Correa, Paulo Faria, Marcelo Fischborn, Chris Kelp, Carlos Sartori, Waldomiro J. Silva Filho, Luca Tateo and Flavio Williges.

## References

- ARMSTRONG, D. M. Is Introspective Knowledge Incorrigible? In: *The philosophical Review*, 72:4 (1963), p. 417-432.
- AXTELL, Guy. Introduction. In: AXTELL, Guy (ed.). *Knowledge, Belief, and Character: Readings in Virtue Epistemology*. Lanham: Rowman & Littlefield, 2000.
- BAEHR, Jason. Character, Reliability and Virtue Epistemology. In: *The Philosophical Quarterly*, 56:223 (2006), p. 193-212.
- BAKER, Gordon; MORRIS, Katherine J. *Descartes' Dualism*. London: Routledge, 1996.
- CASSAM, Quassim. *Self-Knowledge for Humans*. Oxford: Oxford University Press, 2014.
- CODE, Lorraine. Toward a 'Responsibilist' Epistemology. In: *Philosophy and Phenomenological Research*, 45:1 (1984), p. 29-50.
- COMESAÑA, Juan. Reliabilism. In: BERNECKER, Sven; PRITCHARD, Duncan (eds.). *The Routledge Companion to Epistemology*. London: Routledge, 2011. p. 176-186.
- DRETSKE, Fred. Skeptical Doubts About Self-Knowledge. In: BERNECKER, Sven; PRITCHARD, Duncan (eds.). *The Routledge Companion to Epistemology*. London: Routledge, 2011. p. 425-432.
- EVANS, Gareth. *The Varieties of Reference*. Oxford: Oxford University Press, 1982.
- GERTLER, Brie. *Self-knowledge*. London: Routledge, 2011.
- GETTIER, Edmund. Is Justified True Belief Knowledge? In: *Analysis*, 23 (1963), p. 121-123.
- GOLDMAN, Alvin. Epistemic Folkways and Scientific Epistemology. In: *Philosophical Issues*, 3 (1993), p. 271-285.
- GRECO, John. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge, UK: Cambridge University Press, 2010.
- \_\_\_\_\_. The Value Problem. In: BERNECKER, Sven; PRITCHARD, Duncan (eds.). *The Routledge Companion to Epistemology*. London: Routledge, 2011. p. 219-231.
- GRECO, John; TURRI, John. Virtue Epistemology. In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*. URL = <<http://plato.stanford.edu/archives/win2013/entries/epistemology-virtue/>>, accessed Nov. 14, 2014.
- KELP, Christoph. In Defense of Virtue Epistemology. In: *Synthese*, 179 (2011), p. 409-433.
- KIDD, Chad. The Idols of Inner-Sense. In: *Philosophical Studies* (online), October 28, 2014. DOI: [10.1007/s11098-014-0389-6](https://doi.org/10.1007/s11098-014-0389-6)
- KVANVIG, Jonathan L. Virtue Epistemology. In: BERNECKER, Sven; PRITCHARD, Duncan (eds.). *The Routledge Companion to Epistemology*. London: Routledge, 2011. p. 199-207.
- MCGEER, Victoria. The Moral Development of First-Person Authority. In: *European Journal of Philosophy*, 16:1 (2007), p. 81-108.
- MONTMARQUET, James. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield, 1993.

- MORAN, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press, 2001.
- NISBETT, Richard E.; WILSON, Timothy DeCamp. Telling More Than We Can Know: Verbal Reports on Mental Processes. In: *Psychological Review*, 84:3 (1977), p. 231-259.
- PRITCHARD, Duncan; TURRI, John. The Value of Knowledge. In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). URL = <<http://plato.stanford.edu/archives/spr2014/entries/knowledge-value/>>, accessed Nov. 11, 2014.
- RAMSEY, Frank. Knowledge. In: RAMSEY, Frank. *The Foundations of Mathematics and Other Logical Essays*. London: Routledge and Kegan Paul, 1931. p. 110-111.
- RUSSELL, Bertrand. *The Problems of Philosophy*. Oxford: Oxford University Press, 1912.
- SCHMID, Ulla. Where Individuals Meet Society: The Collective Dimensions of Self-Evaluation and Self-Knowledge. In: ZIV, Anita K.; LEHRER, Keith; SCHMID, Hans B. (eds.). *Self-Evaluation: Affective and Social Grounds of Intentionality*. Dordrecht: Springer, 2011. p. 253-273.
- SILVA FILHO, Waldomiro; DAZZANI, Maria Virginia. Self, Sign, Belief, and Agency. In: *Interacções*, 24 (2013), p. 70-92.
- SOSA, Ernest. The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge. In: *Midwest Studies in Philosophy*, 5 (1980), p. 3-25.
- SOSA, Ernest. Consciousness and Self-Knowledge. In: GERTLER, Brie (ed.). *Privileged Access: Philosophical Accounts of Self-Knowledge*. Farnham: Ashgate. p. 253-262, 2003.
- ZAGZEBSKI, Linda. From Reliabilism to Virtue Epistemology. In: AXTELL, Guy (ed.). *Knowledge, Belief and Character: Readings in Virtue Epistemology*. Lanham: Rowman & Littlefield, 2000. (e-book)
- \_\_\_\_\_. *Virtues of the Mind*. Cambridge, UK: Cambridge University Press, 1996.

Recebido em: 04.03.2015  
Aprovado em: 17.06.2015