

***THE ROLE OF EXTERNALISM
IN UNSAFE BUT RATIONAL
REASONING****

**O PAPEL DO EXTERNALISMO
NO PENSAMENTO INSEGURO,
MAS RACIONAL**

Roberto Horácio de Sa Pereira

<https://orcid.org/0000-0002-9117-0755>

robertohsp@gmail.com

*Universidade Federal do Rio de Janeiro, Rio de Janeiro,
Rio de Janeiro, Brasil*

ABSTRACT *Faria's recent publication aims to show that "if anti-individualism is correct, then the obscurity of validity is more widespread than we need to assume for present purposes." In rejecting "a variety of exculpating moves, designed to shield the rationality of the possibly unlucky reasoner," Faria suggests that content externalism can drastically affect our self-understanding as rational beings. I am afraid I have to disagree. Faria overintellectualizes reasoning, always performed from the first-person perspective and is subject to reflexive control by the reasoner. To be sure, content externalism shows that sound reasoning is contingent on the broad environment and therefore does not depend exclusively on the logical abilities of the living being but is a matter of logical luck. Nonetheless, content externalism is no threat to rationality. My diagnosis: By adopting Wittgenstein's strategy of "bringing words back home,"*

* Article submitted on:19/03/2023. Approved on: 15/06/2023.

Faria overlooks the achievements of cognitive science: in its simplest form, “transitive inference,” rationality pervades the animal kingdom.

Keywords *Externalism. Concepts. Rationality. Transitive inference.*

RESUMO *A recente publicação de Faria tem como objetivo demonstrar que, “se o anti-individualismo estiver correto, então a obscuridade da validade é mais ampla do que necessitamos presumir para os objetivos atuais.” Através da rejeição de “uma variedade de movimentos exculpatórios, projetados para blindar a racionalidade de um pensador possivelmente azarado,” Faria sugere que o externalismo de conteúdo pode afetar drasticamente nossa autocompreensão como seres racionais. Temo que tenha que discordar. Faria superintelectualiza o pensamento, sempre performado da perspectiva de primeira pessoa e é sujeito ao controle reflexivo pelo pensador. Para ter certeza, o externalismo de conteúdo mostra que um pensamento saudável é contingente no amplo ambiente e, assim, não depende exclusivamente das habilidades lógicas do ser vivo, mas é uma questão de sorte lógica. Apesar disso, o externalismo de conteúdo não é ameaça à racionalidade. Meu diagnóstico: através da adoção da estratégia de Wittgenstein de “trazer as palavras de volta pra casa”, Faria negligencia as conquistas da ciência cognitiva: na sua forma mais simples, a “inferência transitiva,” a racionalidade pervade o reino animal.*

Palavras-chave *Externalismo. Conceitos. Racionalidade. Inferência transitiva.*

General Overview

Since the 1990s, there has been much debate about rejecting content externalism (anti-individualism) because it is incompatible with the requirement of knowledge about one’s propositional attitudes and contents. The criticism was that it undermines the *a priori* status of authoritative self-knowledge. At first glance, content externalism contradicts the view that one can have prior knowledge about oneself. Content externalism implies that one must explore the external world to become aware of one’s thoughts. It rules out *a priori* self-knowledge and thus undermines the notion that individuals have privileged access to their mental states.

Boghossian argues that content externalism also endangers our logical faculties in cases that Burge calls “slow changes,” in which “the changes take place without disrupting the continuity of one’s life” (Burge, 1988, p. 58). This

critique is based on the transparency Boghossian requires to possess concepts. Under normal conditions, we can determine a priori by introspection whether present, occurring thoughts contain the same or different concepts (Boghossian, 1992, p. 13).

Boghossian's argument is a classic example of *reductio ad absurdum*. If content externalism is true and transparency fails, the thinker would be prone to simple errors in basic reasoning. This conclusion is already a cause for great concern. Even more troubling, if content externalism is true in the slow-switching cases, the subject would be unable to detect and correct these errors in advance. Moreover, it is difficult to determine what empirical investigations the thinker should undertake to correct these errors.

In the literature, we find at least four different positions. Boghossian (1992, 1994, 2011, 2012, 2015) and McKinsey (1991, 2002, 2018) advocate the first position, called "internalist incompatibilism." They claim that the absence of transparency invalidates trivial inferences, implying that externalism is false. The second position, "internalist compatibilism," proposes narrow and broad content to reconcile externalism with transparency. Boghossian argues, "if then, it is also true, that there is an important sense in which [the thinker's] behavior makes sense from his point of view, we would appear to have here an argument for the existence of a level of intentional description which conserves that sense" (1992, p. 28).

Authors such as Burge (1998), Garmendia (2014), Ludlow (1995, 1999), Recanati (2012), Tye (1998), and Sainsbury and Tye (2012, 2015) advocate "externalist compatibilism," which is the third position in response to the incompatibility between the transparency requirement of conceptual possession and content externalism. Camp characterizes this position (externalist compatibilism) as emphasizing the maximization of rationality over truth, implying that reasoning is enthymematic and includes some implicit false premises (Camp, 2002, pp. 38-39). A critical distinction between the "Orwellian" view (Ludlow and Tye) and the "pastiest" view (Burge et al.) is worth noting in cases of slow switching: the former claims that the false premise expresses belief about the past. At the same time, the latter assumes that the false premise expresses belief about the present.

According to Boghossian (2015, p. 79),¹ Sørensen (1998), Williamson (2000), and Faria (2009, 2010, 2021) espoused the fourth and final position, "externalist incompatibilism," which some may consider absurd: it holds that

1 And Faria himself (2021, p. 35).

even basic reasoning is contingent on changes in context. They all claim that reasoning is an unsafe business and reject “a variety of exculpating moves, designed to shield the rationality of the possibly unlucky reasoner against the contingencies of context-shifting” (Faria, 2021, p. 35). On closer inspection, however, while Faria accuses the reasoners of irrationality, Sørensen confines himself to claim they are unlucky under the circumstances. In this paper, I have focused on Faria’s position. Faria himself said, “[T]he point I am wanting to drive home here is that reasoning is bound to be always, to some extent, an unsafe business; also, that this is something we are apt not to understand about the subject matter of logic” (2021, p. 45). Faria has the following: content externalism could drastically affect our understanding of ourselves as rational beings/creatures.

Faria overintellectualizes reasoning as performed from the first-person perspective and subject to reflexive control by the reasoner. By “reflexive control,” Faria means that a premise, whether tacit or explicit, can belong to reasoning only if the reasoner recognizes it from his or her first-person perspective (2021, pp. 44–45). Moreover, “a reasoner cannot coherent be treated as a spectator of her own propositional attitudes and content” (Faria, 2021, p. 30); “the capacity to critically assess one’s own judgment is a constitutive feature of rationality” (Faria, 2021, p. 29). Content externalism shows that sound reasoning is contingent on the broad environment and therefore does not depend exclusively on the logical abilities of the living being but is a matter of logical luck. Nonetheless, content externalism is no threat to rationality. My diagnosis: By adopting Wittgenstein’s strategy of “bringing words back home,” Faria overlooks the achievements of cognitive science: in its simplest form, “transitive inference,” rationality pervades the animal kingdom.

Without further argument, I assume that concepts are representations that enable one to think about the concept. In this sense, one may reasonably think that Oedipus had at least two distinct concepts of his mother: one that might be described as “Mom” and another that arose when he married her, possibly similar to the Greek versions of “Jocasta.” This assumption, however, rules out *a priori* Kripke and Faria’s Millianism, which claims that meaning is exhausted in reference and that Oedipus, therefore, has only one concept of Jocasta. To avoid a *petitio principii*, I will adopt only Evans’ recombination constraint on concept possession. According to this constraint, a person who possesses a singular concept “a” must be able to combine it with any other general concept “F, G, H” that she possesses.

Similarly, a person who possesses the general concept “F” must be able to combine it with any singular concept “a, b, c” that she possesses. For example,

a three-year-old child masters the concept “mommy” if they can recombine this singular concept with any general concept they possess, such as “mommy gives me juice,” “mommy gives me milk,” or “I love mommy.” In addition, they possess the general concept of “milk” if they can combine it with any singular concept they master, such as “mommy gives me milk,” “nanny gives me milk,” or “daddy gives me milk.”

Be that as it may, I will follow Faria’s advice:

[B]ut then the first thing is to get rid of science fiction and follow Wittgenstein’s advice, bringing words back home.

What happens if we detach the examination of the main question from that framework? To begin with, *we will stay on Earth and consider the possibility that the information that the subject lacks about her environment is, after all, available; moreover, the subject would be apprised of it had she only cared to know.*

No “semantics of travel,” then: uncontroversial examples of external individuation of content—beginning, foreseeably enough, with singular thoughts—should suffice for our purposes. (2021, p. 41, emphasis added)

Unlike Faria, I have no objection to “semantic travels,” the famous thought experiments proposed by adherents of content externalism, such as Twin-Earth, Brains in a Vat, and Cosmic Swampbrains. However, I follow Faria’s suggestion and focus my argument on a mundane example: Faria’s acquaintance with a golden retriever from his neighborhood. Following Ryle, I will refer to him as “Fido.”

Faria’s suggestion has the merit of making things clearer. First, without the memory controversies that Burge’s slow-switching cases raise, the mundane cases undeniably show that content externalism threatens the assumption that we can always be introspectively aware of the concepts we primarily harbor rather than the validity of our elementary inferences.

Second, trivial cases allow us to avoid the stalemate in the present literature. For example, in cases where things change slowly, there is no compelling justification for the “Orwellian” view over the “pastiest” view, or vice versa. In this respect, I argue that the debate is indeterminate: there are good reasons to support the “pastiest” view and equally good reasons to support the “presentist” view. Moreover, Burge’s anaphoric account of memory storage, ingenious as it may be, raises serious questions in the philosophy of memory. Bernecker (1998), for example, states that “the transfer of contents and concepts over time may be a sufficient condition for memory, but it does not necessarily have to be” (Bernecker, 1998, p. 241). Once again, we are confronted with a gray area of indeterminacy, as the case extrapolates the actual world and natural environment in which memory developed. Be that as it may, Wittgenstein’s

methodological advice always has the same fatal consequence: it blinds the researcher to the achievements of cognitive science in general.

I organize this paper as follows. In the section following this introduction, I will examine Boghossian's constraint on conceptual sameness transparency, which holds that we can know a priori, and even by introspection alone, that we are using the same concept in elementary inferences. I will argue in this section that Boghossian's constraint on conceptual sameness is too intellectual: simple-minded beings need not know introspectively a priori whether they are using the same concept when they draw simple elementary inferences. But Faria's claim that "attributing rationality to an agent amounts necessarily to ascribing her a *prima facie* privileged access to the contents of her own propositional attitudes" (2021, p. 29) also overintellectualizes reasoning since simple-minded reasoners have no privileged access to their contents.

In the second section, I will examine Boghossian's constraint on the transparency of conceptual difference in more detail, which holds that we can know a priori and by introspection alone when using two or more concepts in elementary reasoning. I will argue in this section that Boghossian's constraint on conceptual difference is over-intellectualized: simple-minded beings know a priori and, by introspection, whether they use more than one concept. Again, Faria's claim that "attributing rationality to an agent amounts necessarily to ascribing her a *prima facie* privileged access to the contents of her own propositional attitudes" (Faria, 2021, p. 29) also overintellectualizes reasoning.

In the last part, I draw my conclusions: Simple-minded beings do not know whether they use the same or different concepts in the premises of their transitive inferences. Moreover, their transitive inferences are beyond their reflexive control. Content externalism shows how reasoning becomes an unsafe business when the reasoner tacitly assumes false conceptual identities and differences due to changes in the environmental context. Nothing, however, poses a threat to the creature's rationality.

1 - Transparency of sameness

Since the 1990s, Boghossian has launched a persistent attack on content externalism. According to his main objection, content externalism contradicts his content or conceptual transparency claim. In Boghossian's own words:

[Content externalism] is inconsistent with the thesis that our thought contents are epistemically transparent to us [...] This is true in a sense that falsifies another important and traditionally held view—that we can detect a priori whether our inferences are logically valid or not. (1992, p. 13)

I will explain the dialectic in detail, step by step. First, the general statement about content transparency as a constraint is as follows:

- **Introspective knowledge of comparative content.** If our capacity for introspection functions normally, we can know *a priori* by introspection whether two present, occurring thoughts use the same or different concepts.

Boghossian's constraint restates the old idea that mastering a concept requires distinguishing it from others. For example, if a three-year-old child has a concept of "mommy," he must be able to distinguish it from his concept of "nanny." On closer inspection, however, it becomes clear that Boghossian's transparency constraint returns to the time-honored tradition that began with Descartes. Descartes postulated that he could doubt anything; he could reasonably assume that most of his beliefs could be false and even that he could not know the existence of a world outside his mind. Nevertheless, beginning with his *Second Meditation*, Descartes was convinced that he would never be wrong when considering the concept of God or the concept of an evil genius.

If we leave tradition aside, the general claim divides into two parts: the transparency of sameness and the transparency of difference.

- **Transparency of sameness.** If S's capacity for introspection is functioning normally, then if S happens to think two thoughts that use the same concepts, he can introspectively recognize this fact.

Suppose S's capacity for introspection is functioning adequately. In this case, S can introspectively recognize this identity when he has two thoughts involving the same concept. In any case, content externalism contradicts what Boghossian calls the transparency of sameness. Above all, this raises the problem that rejecting the transparency of sameness leaves us vulnerable to the fallacy of "univocation" (Sørensen, 1998, p. 326) in even the simplest inferences, where the thinker considers the same concept as if it were multiple.

Let us examine Faria's story if we heed Wittgenstein's advice to "bring the words back home" (Faria, 2021, p. 41). While walking home one afternoon, Faria paused and spotted a friendly-looking golden retriever in his neighbor's front yard. Faria took some time to pet him. After a while, Faria started playing ball with him and rewarded Fido with some cookies. As he headed home, Faria thought:

1. Fido is a friendly dog.

Faria saw a stunning golden retriever sprinting through the backyard a few days later. Faria tries to get the dog's attention by whistling and clapping his hands, but to no avail. The animal continues to run around, barking at every

tree it encounters. When Faria returns home, he remembers that his neighbor is a breeder of golden retrievers, and he comes up with the following thought:

2. This dog (the same Fido) is not friendly.

Now Faria's neighbor has only one golden retriever, Fido. Consequently, "Fido" refers to a single golden retriever. Is there a Fregean solution to Faria's predicament? The Fregean solution is that Faria (the reasoner) associates two different modes of presentation (two "senses" or "modes of representation") with the same name, "Fido," referring to the same dog: FIDO. In statement 1, "Fido" stands for FIDO as a friendly dog, while in statement 2, "Fido" stands for the same FIDO as a highly agitated animal. Faria (the philosopher) hastily dismisses the Frege-like solution, arguing that "the difference, most certainly here anyway, 'ain't in the head." If there are two Fregean senses at play here, they are *de re* senses, their identity partially fixed by the subject's non-representational relation to his environment" (2021, p. 45).

But even if the "senses" or "modes of presentation" are "*de re*" and not "*de dicto*," this does not mean that they "ain't in the head," as Faria claims. The term "*de re*" sense typically refers to determining a reference through non-representational relations to the environment. The "*de re*" sense exists only when the reference exists; hence, they are entirely mind-independent. Nevertheless, neo-Fregeans may have a broader understanding of the "*de re*" sense. It is not only a way of presenting reference through non-representational relations to the environment but also a way of representing reference in a relational, nondescriptive way. In this latter sense, the "*de re*" senses are what Bach (1987) calls "*de re*" modes or ways of presenting reference. What determines reference is not the satisfaction of an identifying condition but acquaintance relations (see Bach, 1987, p.12). The point is that "*de re*" modes of presentation are not dependent on the existence of their referents and, as such, can be "in the head" as "mental files." Recanati's framework illustrates how easily we arrive at a solution consistent with content externalism:

Paderewski cases are cases where a subject associates two distinct files with a single name. Inverse Paderewski cases are cases in which there are two names, but the subject associates them with a single file. (2012, p.116)

According to Recanati's framework of mental files, concepts as mental files undergo various processes, such as splitting, merging, and linking. With this in mind, the simplest cognitive explanation for Faria's predicament is that Faria encounters Fido in two situations where the dog behaves differently. This creates two unconnected mental files in Faria's mind that refer to the same dog, Fido. This would explain why Faria believes that Fido in situation 1 is not the

same as in situation 2. However, for reasons unknown (at least to me), Faria (the philosopher) wants to resist this neo-Fregean solution to Faria's (the reasoner) predicament. Faria adopts Kripke's Millianism without further argument as if this view were the only semantics compatible with content externalism. Faria (the reasoner) has only one concept of FIDO because meaning is a reference, and there is only one dog in the story: Fido.

Boghossian's view is correct in that content externalism is inconsistent with his claim of transparency of sameness. Since Putnam's (1975) seminal work on the linguistic/cognitive division of labor, it has been widely recognized that one cannot introspectively distinguish between different concepts. Burge's (1979) famous example involving an orthopedist and a patient illustrates this point well. The patient complained of arthritis in his thigh, whereupon the doctor explained that arthritis was a joint disease and that the patient had "gout." The patient then corrected himself, showing respect for the orthopedist's expertise. The question is: Can the patient introspectively distinguish between arthritis and gout? Does the patient introspectively know *a priori* whether he has only one concept of arthritis or more than one? No, of course not! Since Putnam's division of linguistic labor, no one has accepted the claim that one can possess one concept only if one can distinguish it from another.

Faria's introspective abilities work correctly. However, he has to recognize *a priori* introspectively that he uses the same concept, FIDO, in both thoughts, which he cannot do. Consequently, Faria falls prey to the "fallacy of univocation" (Sørensen, 1998, p. 326) by treating the same concept as if it were more than one concept:

4. Fido is a friendly dog & Fido is not a friendly dog.

If we take "FD" to stand for the predicate "friendly dog" and "b" to stand for FIDO, the blatant contradiction takes the following form:

1. FDb.
2. ~FDb.

4. FDb & ~FDb.

The thinker (Faria) could have easily avoided the contradiction if he had fulfilled Boghossian's requirement. Since he knew *a priori* that he was using only one concept in both ideas, he would not have upheld 2 and 4. Unfortunately, Faria could not introspectively recognize *a priori* that he was using the same concept, FIDO, in both thoughts. As a result, his reasoning culminates in a glaring contradiction.

Boghossian never considered that an individual might fall prey to the fallacy of univocation in such basic reasoning. Boghossian's argument against content externalism takes the form of a *reductio*:

- Content externalism is true.
- There is no transparency of conceptual identity.
- Everyone would be susceptible to the fallacy of "univocation."
- Premise 2 is completely implausible, and Premise 4 is absurd.

-
- Content externalism is false.

Faria takes content externalism to be true and reaches the astonishing conclusion that everyone can fall prey to simple fallacies:

- Content externalism is true.
- There is no transparency of conceptual identity.
- Everyone would be susceptible to the fallacy of "univocation."
- Premise 2 does not hold.

-
- Everyone is susceptible to the fallacy of univocation.

The implications of content externalism suggest that elementary inference is an unsafe business. In summary, Faria (2021, p. 41) notes: "If anti-individualism is correct, the opacity of validity will be more pervasive than we must assume for the aims in question" (Faria, 2021, p. 41, emphasis added). Faria is willing to sacrifice the reasoner's rationality in favor of his true beliefs; instead of emphasizing the opacity of concepts, he posits the opacity of validity. However, Boghossian takes a different approach; he sacrifices content externalism in favor of rationality. What is a *modus tollens* for Boghossian is a *modus ponens* for Faria:

- ❖ If content externalism is correct, everyone can fall prey to the fallacy of univocation.
- ❖ Not everyone can fall prey to the fallacy of univocation.

-
- ❖ Ergo, content externalism is false (Boghossian).

Faria argues as follows (*modus ponens*):

- If content externalism is correct, everyone can fall prey to the fallacy of univocation.
- Content externalism is right.

-
- Ergo, everyone can fall prey to the fallacy of univocation.

If we prioritize truth over rationality, we risk failing to understand the thinker's thought process. Indeed, it is difficult to understand an argument that ends in such a glaring contradiction:

4. Fido is a friendly dog & Fido is not a friendly dog.

If we are to hold our commitment to externalism, then we must assume that the argument is an enthymeme containing a tacitly false premise, such as:

3. The FIDO thought of in 1 is different from the FIDO thought of in 2.

Rather than falling for this notorious fallacy, the most plausible explanation for what the thinker has in mind is unsound but valid reasoning:

1. FDb.

2. ~FDb.

3. b in 1 (b1) \neq b in 2 (b2) (the tacit false premise).

4. FDb1 & ~FDb2.

Faria vehemently rejects the idea that 3 should be considered a tacit premise in enthymematic reasoning. He argues that this inference is a kind of exculpation “designed to shield the rationality of the possibly unlucky reasoner against the contingencies of context-shifting” (Faria, 2021, p. 35). Logical thinking and reasoning can be risky, even for the simplest inferences. Nonetheless, one must reflect on Faria's reason for rejecting 3 as a tacit premise in enthymematic reasoning:

An identity statement will be taken as a premise, whether tacit or explicit, only if there is a difference recognizable from the first-person perspective (in other words, by reflection alone) between the ways the object is thought of in each pair of premises. That is why, should I reason from “The dog I met on Tuesday is very friendly” and “The dog I met on Friday is very restless,” I would be helping myself to the identity “The dog I met on Tuesday = The dog I met on Friday” as a premise. (2021, pp. 44-45, emphasis added)

Faria (the philosopher) claims that an identity statement can occur as implicit premise 3 in a simple argument from 1 to 4 only if the arguer can recognize the difference between Fido in 1 and Fido in 2 from his first-person perspective. This requires the reflexive ability of the arguer to consider premises 1 and 2 in turn and to ask himself whether the dog in 1 is the same as in 2. But is this a good reason to support the claim that the validity of elementary reasoning is as opaque as the concepts involved?

Faria's claim overintellectualizes simple reasoning. With such a stringent criterion, it is no wonder that any rational person may be irrational. Why must a creature reflexively infer 1 from 2 and wonder whether the dog in 1 is the same as the dog in 2 for 3 to appear as a premise in his argument?

This is overintellectualization. To illustrate, three-year-old children can easily understand many concepts and apply them confidently and repeatedly in different ways to think about the world. For example, a three-year-old child might think, “Mommy is sweet, Mommy loves Bobby.”

Imagine a three-year-old child calling out to his mother twice. First, the mother responds with a comforting hug. The next day, however, the child’s mother is busy and cannot respond similarly. As a result, the child may think his nanny is responding to his calls for help. In light of all this, the three-year-old child might consider the following:

5. Mommy is a kind caregiver.

6. Mommy (nanny) is not a friendly caregiver.

8. Ergo, Mommy is a friendly caregiver & Mommy is not a friendly caregiver.

To avoid unwanted contradiction, we usually assign a tacit false premise to the child as follows:

7. Mommy in 5 is not Mommy in 6.

Can this three-year-old child reflexively infer from 5 to 8 and wonder if the referent in 5 is the same as in 6 to falsely assume that his referent in 5 is the same as in 6? “The capacity to critically assess one’s own judgment is a constitutive feature of rationality” (Faria, 2021, p. 29). This is too demanding. Because of this, we use the attribution of a presupposed conceptual difference in the third person as a tacit premise in an enthymeme to keep the inference from leading to a contradiction or irrationality:

4. FDb & ~FDb.

Consequently, statement 3 is a third tacit premise (in enthymematic reasoning) as an unreflective belief. Early humans thought and talked about the world long before they began to understand the reason for their conclusions and arguments. Thinking within the framework of content externalism may be an unsafe business. However, the individual who engages in such reasoning is not necessarily irrational. But what is the cognitive explanation for the fact that the baby assumes 7?

For the sake of argument, let us assume that Kripke’s Millianism is the only game in town. Faria meets Fido for the first time, and the dog behaves very friendly toward him. This encounter generates a single concept, FIDO. If Millianism is true, Faria has no additional concept of Fido. However, when Faria encounters Fido for the second time, and the dog behaves quite differently, Faria’s concept splits into Faria’s mind as if there were two concepts of two dogs under the same “mental file.” Therefore, Faria (the reasoner) unreflectively

believes that the Fido in 1 is not the same as the Fido in 2. This happens to the reasoner and is not something he reflexively performs, which contradicts Faria's claim that "a reasoner cannot coherently be treated as a spectator of her propositional attitudes and content" (2021, p. 30). However, this is not to say that what moves Faria is just blind causality without reason. It only means the thinker has no reflexive control, relieving Faria of irrationality. Faria's reasoning is thoroughly rational but infelicitous and unlucky.

There are other reasons to distrust Faria's view that content externalism could drastically affect our self-understanding as rational thinkers. In contrast to Faria's top-down approach, a bottom-up perspective suggests that, in nature, it is not fallacious inferences (see Faria, 2021, p. 60) but rather rationality that is pervasive. Throughout the animal kingdom (including insects), where there is social cooperation, there is cognition and thus "transitive inferences." Tibbetts, Agudelo, Pandit, & Riojas (2019) have convincingly shown, for example, that the social wasp (the so-called paper wasp) is capable of a type of transitive reasoning-like behavior:

Transitive inference (TI) is a form of logical reasoning that involves using known relationships to infer unknown relationships ($A > B$; $B > C$; then $A > C$). TI has been found in a wide range of vertebrates but not in insects. Here, we test whether *Polistes dominula* and *Polistes metricus* paper wasps can solve a TI problem. Wasps were trained to discriminate between five elements in series (A0B2, B0C2, C0D2, D0E2), then tested on novel, untrained pairs (B versus D). Consistent with TI, wasps chose B more frequently than D. Wasps organized the trained stimuli into an implicit hierarchy and used TI to choose between untrained pairs. Species that form social hierarchies like *Polistes* may be predisposed to spontaneously organize information along a common underlying dimension. This work contributes to a growing body of evidence that the miniature nervous system of insects does not limit. (2019, p. 1)

When Fido sees Faria coming home in the afternoon, he spies him from the neighbor's garden. Faria's affectionate entreaties get Fido to approach him, and they play ball. Faria rewards Fido with some cookies. Fido forms the singular concept of FARIA without engaging in the reflexive activities of comparison, separation, and abstraction. Fido acquires the concept of FARIA through direct contact with Faria, resulting in a state of thought in his mind. I am now assuming that something like a thought state passes through Fido's mind, roughly speaking:

8. Faria is a nice guy.

A few days later, something changed. When Faria came home, and Fido saw him in the dark, he started barking and behaving aggressively toward him. It is safe to assume that a different state of mind arose in Fido's head, something like this:

9. Faria is not a nice fellow but an intruder.

Is it reasonable to assume that Fido is susceptible to the “fallacy of univocation” to explain his initially friendly and later aggressive behavior toward Faria in the following argument?

8. Faria is a nice guy.

9. Faria is not a nice guy.

11. Faria is a nice guy & Faria is not a nice guy.

The answer is no. Fido is not irrational. Once again, rationality is pervasive. The only way to understand Fido’s contradictory behavior towards the same person, namely Faria, is to impute an implicit false premise to him (his enthymematic reasoning):

10. Faria (thought of in 8) is not Faria (thought of in 9).

Let us assume Kripke and Faria’s Millianism for the sake of argument again. In the first encounter, Fido behaves quite kindly toward Faria, as a professor returning home. In contrast, in the second encounter, Fido behaves aggressively, assuming that Faria is an intruder. In this case, Fido behaves as if he has two singular concepts, although, if Millianism is true, he has only one singular concept, FARIA. So Fido’s “transitive inference” is rough as follows:

8. Faria is a nice guy.

9. Faria is not a nice guy.

10. Faria (thought of in 8) is not Faria (thought of in 9) (the tacit false premise).

11. Faria in 8 is a nice guy & Faria in 9 is not.

Fido’s belief takes the form of tacit premise 10, namely that the Faria mentioned in 8 is not the same Faria referred to in 9. Now the question is again, what accounts for the fact that Fido assumes 10? If Kripke and Faria’s Millianism is true, Fido has only one concept of FARIA. Since it is a reference, Faria is the only human animal in the story. Yet, Fido’s experience with Faria in two different situations splits Fido’s concept of Faria as if he has two different concepts of two people.

2 - Transparency of difference

The second part of the content-transparency requirement is as follows:

- **Transparency of difference.** If S’s capacity for introspection works normally, then if he is currently thinking two thoughts that use different concepts, he can introspectively recognize this fact.

If the ability to introspect is functioning appropriately, one should be able to recognize a priori when thoughts use different concepts—content externalism conflicts with this requirement. The major concern is that if we abandon the clarity of conceptual distinctions, we may fall prey to the fallacy of equivocation in basic reasoning or inference, i.e., the thinker takes different concepts as if they were the same thing.

Following Wittgenstein’s call to “bring the words back home” (Faria, 2021, p. 41), I consider not only Faria’s golden retriever case but now the supposed fallacy of equivocation. Let me recapitulate. When Faria came home, he encountered a golden retriever named “Fido” playing around in his neighbor’s front yard. The dog was adorable, and Faria was happy to pet him. He thought fondly back on his encounter with Fido:

12. Fido was friendly to me yesterday.

A few days later, a similar situation occurs when Faria encounters another golden retriever, also named “Fido,” walking around (let us assume the dog has his name on his collar). Faria stops again, hoping to get the dog’s attention, and as in the first encounter, the dog behaves friendly. But this time, a different thought crosses Faria’s mind:

13. Fido is friendly to me today.

From 12 and 13, Faria concludes:

15. Fido is a friendly dog.

The problem is that the name “Fido” is no longer associated with a single golden retriever but with two indistinguishable ones, for example, FIDO1 and FIDO2. Consequently, Faria’s elementary reasoning falls prey to the fallacy of equivocation (when one uses a term with two or more meanings in the same argument).

We face the dilemma of prioritizing a subject’s rationality or actual beliefs. Davidson’s golden rule states that, given all relevant information, it is always better to prioritize rationality over truth because we make daily mistakes. However, we rarely fall for simple fallacies in transitive inferences. By giving preference to truth over rationality, we limit our understanding of the thinker. Davidson’s golden rule, therefore, strongly recommends the assumption that statement 14 is an implicit false premise of enthymematic reasoning:

14. FIDO in 12 is FIDO in 13.

If we assume that “N” stands for the predicate “beautiful” and “b” for FIDO, the most plausible explanation for what Faria has in mind is the following unsound but valid argument:

12. Nb at t1.

13. Nb at t2

14. b at 1 = b at 2.

15. Nb.

Once again, Faria rejects this strategy to escape the conclusion that argumentation is a precarious venture. To reject 14 as a tacit, implicit premise, Faria argues:

An identity statement will be taken as a premise, whether tacit or explicit, only if there is a difference recognizable from the first-person perspective (in other words, by reflection alone) between the ways the object is thought of in each pair of premises. That is why should I reason from “The dog I met on Tuesday is amiable” and “The dog I met on Friday is very restless,” I would be helping myself to the identity “The dog I met on Tuesday = The dog I met on Friday” as a premise. (2021, pp. 44-45, emphasis added)

My diagnosis remains the same: overintellectualism. Complex thinking can mean recognizing the identity of concepts through reflection from the first-person perspective when the thinker considers one premise after another and wonders if he is using different concepts. In “transitive inference,” pervasive in the animal kingdom, however, one can assume that the creature accepts the concept of identity without reflection from his first-person perspective. How can we account for the fact that Faria assumes belief 14? Could a Fregean offer a solution to this puzzle?

We have a neo-Fregean, at least insofar as we assume the existence of “*de re*” modes of presentation à la Bach (1987). As we have seen, neo-Fregeans may have a broader understanding of the “*de re*” sense, namely as a way of presenting reference through non-representational relations to the environment. These are also nondescriptive ways of presenting references, like mental files in the reasoner’s head. The simplest explanation of Faria’s predicament is as follows: Faria’s (the reasoner) acquaintance with qualitatively similar dogs that behave similarly creates a single mental file in Faria’s mind that equivocally refers to two qualitatively similar but numerically different dogs. This explains why Faria assumes 14 and concludes 15.

But again, for reasons unknown, Faria (the philosopher) wants to resist this neo-Fregean solution to Faria’s (the reasoner) predicament. Without further argument, Faria adopts Kripke’s Millianism as if this view were the only semantics compatible with content externalism. Faria (the reasoner) has only one concept of FIDO because meaning is a reference, and there is only one dog in the story: Fido. In a Millianist framework, there is no way to excuse Faria for his equivocation. Faria not only bites the bullet but swallows it afterward.

Let us imagine ourselves in the place of one of Faria’s golden retrievers, encountering Faria in the afternoon. At Faria’s friendly and affectionate invitation, Fido approaches him and behaves cordially. They play ball, and

Faria rewards Fido with some cookies. Because of his acquaintance with Professor Faria, Fido forms the singular concept of FARIA. A thought-like state arises in Fido's mind:

16. Faria was nice to me today.

Now suppose that Faria has a twin who comes to visit. Fido sees Twin Faria in the garden, and at his request, he approaches Twin Faria and behaves just as amicably. They play ball together, and twin Faria gives the dog some cookies. Therefore, we can rightly attribute the following thought to Fido:

17. Faria was nice to me another day.

The next time Fido meets Faria (not his twin), he wags his tail and shows great affection towards him. This behavior, combined with thoughts 16 and 17, leads us to conclude that Fido thinks as follows:

19. Faria is a nice guy.

Fido argues as follows:

16. Faria was nice to me yesterday.

17. Faria was nice to me today.

19. Faria is a nice guy.

The problem is that the concept "Faria" does not refer to a single person, Faria, but to two identical people (at least in all aspects relevant to a dog), namely Faria and twin Faria. Does this mean that Fido behaves irrationally?

Of course, this makes no sense in the light of cognitive science. The only plausible assumption is that Fido has mistakenly recognized twin Faria as Faria, whom he saw the day before. Fido's transitive inference tacitly assumes a false premise, namely that:

18. Twin Faria is Professor Faria.

Therefore, Fido reasons along the following lines:

16. Faria was nice to me yesterday.

17. Twin Faria was nice to me today.

18. Twin Faria is Professor Faria (the tacitly false premise).

19. Faria is a nice guy.

How is it to be explained that Fido tacitly assumes 18 and, in this way, avoids falling prey to the fallacy of equivocation? The cognitive answer is obvious. The fact that Fido does not distinguish between Faria and his twin creates a single concept/file in his mind. This makes his transitive inference unsound but not irrational.

If you are skeptical about animal transitive inference, consider a three-year-old child interacting with his mommy and nanny.² When the three-year-old child cries on Monday, the mother sings lullabies to the child. The child enters a thought-like state:

20. Mommy was nice to me today.

Now suppose that the three-year-old child cries again on Friday. This time, however, it is not Mommy who comes to the rescue but his nanny. The nanny takes the child in her arms, who thinks:

21. Mommy was nice to me another day.

Thoughts 20 and 21 entitle us to attribute the following conclusion to the baby:

22. Mommy loves the baby (me).

The three-year-old child reasons as follows:

17. Mommy was nice to me on Monday.

18. Mommy was nice to me on Friday.

20. Mommy loves the baby (me).

The concept of “mommy” in this context is not limited to one person but refers to two caregivers: the baby’s mother and the nanny. Could the three-year-old child be making the fallacy of equivocation in his early logical reasoning? Of course, this makes no sense. The only plausible assumption is that the baby has mistakenly recognized his nanny as his mother. The three-year-old child thus tacitly assumes a false premise, namely:

19. The nanny is Mommy.

Therefore, the baby reasons along the following lines:

17. Mommy was nice to me on Monday.

18. The nanny was friendly to me on Friday.

19. The nanny is Mommy (the false tacit premise).

20. Mommy loves the baby (me).

Faria complains that the false identity 15 can only be tacitly or explicitly accepted as a premise when the thinker reflexively reconsiders each premise from his first-person perspective and wonders if he is succumbing to a mistake:

21. Does “Mommy” in 17 refer to the same person as “nanny” in 18?

2 Those who remain unconvinced that dogs can intuitively follow rules of inference will be impressed by the case of Chase, a Border Collie who learned the names of 1022 toys, regardless of shape and color. To test the dog’s reasoning abilities, the researcher (with the toys hidden behind his back) asked Chase to find a toy whose name the canine was unfamiliar with: “meow.” In a single attempt, the Border Collie could intuitively identify the toy using the propositional calculus elimination rule, a remarkable feat captured on video: https://www.youtube.com/watch?v=lp_uVTWfXyl.

The answer implies that a person can only be a rational thinker if he has a higher-order ability to meta-represent the concepts he uses in his reasoning. Is that not too intellectual?

Conclusion

Faria assumes that content externalism can undermine our most basic reasoning processes, e.g., through the fallacies of univocation and equivocation. However, content externalism does not affect the rationality of basic, transitive inferences but rather our full understanding of the concepts involved in those inferences. Faria's view makes an intellectualist assumption regarding reasoning. At the most basic level, transitive inferences are not within the control of the thinker.

References

- ANDREWS, K. "How to Study Animal Minds". Cambridge University Press, 2020.
- BACH, K. "Thought and reference". New York: Clarendon Press, 1987.
- BERNECKER, S. "Self-Knowledge and Closure." In: P. Ludlow, N. Martin (eds.). *Externalism and Self-Knowledge*. Stanford: CSLI Publications, 1998. pp. 333-347.
- BOGHOSSIAN, P. "Externalism and inference." *Philosophical Issues*, 2, 1992, pp. 11-28. <https://doi.org/10.2307/1522852>.
- _____. "Further Thoughts on the Transparency of mental content." In: S. C. Goldberg (ed.). *Externalism, self-knowledge, and skepticism, new essays*. Cambridge: Cambridge University Press, 2015. pp. 95-97.
- _____. "The transparency of mental content revisited." *Philosophical Studies*, Vol. 155, Nr. 3, 2011, pp. 457-465. <https://doi.org/10.1007/s11098-010-9611-3>.
- _____. "The transparency of mental content." *Philosophical Perspectives*, 8, 1994, pp. 33-50. <https://doi.org/10.2307/2214162>.
- _____. "What is inference?." *Philosophical Studies*, Vol. 169, Nr. 1, 2012, pp. 1-18. <https://doi.org/10.1007/s11098-012-9903-x>.
- BURGE, T. "Individualism and Self-Knowledge." *Journal of Philosophy*, Vol. 85, Nr. 7, 1988, pp. 339-362.
- _____. "Individualism and the mental." *Midwest Studies in Philosophy*, Vol. 4, Nr. 1, 1979, pp. 73-122.
- _____. "Memory and self-knowledge." In: P. Ludlow, N. Martin (eds.). *Externalism and self-knowledge*. CSLI, 1998. pp. 111-127.
- CAMP JR., J. L. "Confusion: a Study in the Theory of Knowledge". Cambridge, MA: Harvard University Press, 2002.
- FARIA, P. "Memory as acquaintance with the past: Some Lessons from Russell, 1912-1914". *Kriterion: Revista de Filosofia*, Vol. 51, Nr. 121, 2010, pp. 149-172.

- _____. “Unsafe reasoning: A survey.” *Dois Pontos*, Vol. 6, Nr. 2, 2009, pp. 185-220.
- _____. *Time, Thought, and Vulnerability. An Inquiry in Cognitive Dynamics. The Juan Larreta Lectures.* Ciudad Autónoma de Buenos Aires: SADAF, 2021.
- GARMENDIA, M. E. “Externalism, rational explanation, identity premises.” *Teorema*, Vol. XXXIII, Nr. 3, 2014, pp. 31-48.
- KAPLAN, D. “Externalism and self-knowledge.” *In: The Oxford Handbook of Epistemology.* Oxford University Press, 2010. pp. 206-225.
- LUDLOW, P. “Externalism, Self-Knowledge, and the Prevalence of Slow Switching.” *Analysis*, Vol. 55, Nr. 1, 1995, pp. 45-49.
- _____. “First-person authority and memory.” *In: M. De Caro (ed.). Interpretations and Causes: New Perspectives on Donald Davidson’s Philosophy.* Kluwer Academic Publishers, 1999.
- MCKINSEY, M. “Anti-individualism and privileged access.” *Analysis*, Vol. 51, Nr. 1, 1991, pp. 9-16.
- _____. “Forms of externalism and privileged access.” *Philosophical Perspectives*, 16(s16), 2002, pp. 199-224.
- _____. “Skepticism and content externalism.” *Stanford Encyclopedia of Philosophy*, 2018.
- PUTNAM, H. H. “The meaning of ‘meaning’.” *Minnesota Studies in the Philosophy of Science*, 7, 1975, pp. 131-193.
- RECANATI, F. *Mental files.* Oxford, UK: Oxford University Press, 2012.
- SAINSBURY, R. M., TYE, M. “Counting concepts: Responses to Paul Boghossian.” *In: S. C. Goldberg (ed.). Externalism, Self-knowledge and skepticism, new essays.* Cambridge: Cambridge University Press, 2015. pp. 113-119. <https://doi.org/10.1017/CBO9781107478152>.
- SAINSBURY, R. M., TYE, M. *Seven puzzles of thought.* New York: Oxford University Press, 2012.
- SEARLE, J. R. “Consciousness, unconsciousness and intentionality.” *Philosophical Issues*, 1, 1991, pp. 45-66.
- SØRENSEN, R. A. “Logical luck.” *The Philosophical Quarterly*, Vol. 48, Nr. 192, 1998, pp. 319-334. <https://doi.org/10.1111/1467-9213>.
- TIBBETTS, E. A. *et al.* “Transitive inference in *Polistes* paper wasps.” *Biology Letters*, Vol. 15, Nr. 5, 2019.
- TYE, M. “Externalism and memory.” *Proceedings of the Aristotelian Society, Supplementary Volume*, Vol. 72, Nr. 1, 1998, pp. 77-94.
- WILLIAMSON, T. “Knowledge and Its Limits”. Oxford: Oxford University Press, 2000.