

**teorema**

Vol. XXXIII/1, 2014, pp. 00-00

ISSN: 0210-1602

[BIBLID 0210-1602 (2014) 33:3; pp. 00-00]

## **Know Thyself: A Tale of Two Theses and Two Theories**

Neftalí Villanueva Fernández

### RESUMEN

El propósito de este artículo es defender un análisis expresivista del significado veritativo-funcional de las adscripciones de auto-conocimiento. En primer lugar, presentamos las dos tesis sobre el auto-conocimiento en las que nos centraremos: autoridad fuerte y autoridad presunta. A continuación, ofrecemos un contraste entre el género de expresivismo que defendemos y sus competidores más importantes, agrupados bajo la etiqueta 'descriptivismo'. Mediante la introducción de diversos casos presentamos algunos de los datos que necesita acomodar una teoría para analizar con éxito las adscripciones de actitud proposicional que aquí nos interesan: las adscripciones de auto-conocimiento ni tienen siempre autoridad ni carecen siempre de ella. Finalmente, argumentaremos que el género de expresivismo que apoyamos ofrece el análisis más parsimonioso de esas atribuciones y que, por ello, debe ser considerado como más adecuado que sus alternativas.

PALABRAS CLAVE: *auto-conocimiento, autoridad, expresivismo.*

### ABSTRACT

The purpose of this paper is to defend an expressivist analysis for the truth-conditional meaning of self-knowledge ascriptions. First, we present the two theses about self-knowledge on which we shall focus, Strong Authority and Presumptive Authority; thereafter, we offer a contrast between the kind of expressivism which we will be advancing and some of its major competitors, grouped under the label 'descriptivism'. Through the introduction of several different cases, we present some of the data that a theory needs to accommodate in order to successfully analyze attitudinal avowals: self-knowledge ascriptions are not always authoritative, or non-authoritative, in any sense. Finally, we will argue that the kind of expressivism that we favor gives the most parsimonious analysis of these attributions, and thus should be considered better than its alternatives.

KEYWORDS: *Self-Knowledge, Authority, Expressivism.*

### I. SELF-KNOWLEDGE, NESTED ATTITUDINAL AVOWALS, AND THE TWO THESES

Even though we live tens of centuries apart from the kind of society where an unexamined life was deemed to be not worth living, we certainly

are worried about ourselves. And not just about the human species generically, but also about our specific features, including those that do not necessarily make us unique, but exert a strong influence on the way we live our lives. When things get rough, we often have the sudden urge to understand why we tend to act or feel in a peculiar way. Some undertake therapy, others talk extensively to their friends about themselves, take long solitary walks, or start practicing endurance sports to find a moment in their day to reflect upon themselves. To know oneself seems to be crucial at times. You do not want your reflection upon your best options to be jeopardized by unattended character traits; when you find yourself on a bumpy road in life, you do not want your decisions to be completely determined outside the sphere of your rational control. In this paper we explore certain features of the truth-conditional meaning of self-knowledge ascriptions, the linguistic material that we use to discuss our personal epistemic achievements.

It is part of the Wittgensteinian heritage that ‘phenomenal avowals’, such as ‘my ankle is sore,’ or ‘I have a headache’ exhibit certain particular features: they are taken to be *groundless*, *transparent*, *authoritative*, and *immune to error through misidentification*. They are *groundless* because the speaker who states ‘I have a headache’ cannot be meaningfully required to produce reasons that back up her claim. These avowals are also transparent<sup>1</sup> because it makes no sense to wonder whether the speaker knows that her head aches, *authoritative* because a speaker sincerely uttering ‘I have a headache’ is enough to guarantee the truth of her claim. Also, these avowals are *immune to error through misidentification* because the speaker who claims to have a sore ankle cannot mistake the subject of the experience for that of anybody else’s.<sup>2</sup>

Phenomenal avowals, crucial as they are for a lot of our everyday interactions, do not play a big role in the traditional quest for self-knowledge. If a person who has successfully undertaken Socrates’ directive is one who is well aware of her powers and limitations, able to anticipate and modulate her reactions to her own advantage when confronted with new situations, phenomenal avowals only help to characterize to a certain extent the kind of abilities that this person possesses. We also need to take into consideration a different type of avowals, ‘attitudinal avowals’, such as ‘I think it’s going to rain tonight,’ and, ‘I remember what you said that day,’ and similar. According to Wright [Wright (1998), p. 15], ‘the psychological characteristics, processes and states which they concern are partially individuated by the propositional content, or intentional direction, which informs them’; they are mental-state ascriptions with the form of propositional attitudes. Even more helpful will be what some have called ‘second-order ascriptions’ of mental states (see, for example, Burge (1988), p. 72, Medina (2006), p. 579), as ‘I think that I think it’s going to rain tonight,’ or ‘I know that I remember what you said that day’. Some of the abilities that a person following the Socratic directive can acquire are generalizations over these ascriptions. Someone

who knows herself can know, for example, that every time that she visits her mother she feels that her life amounts to nothing. If she knows that, she can exercise some control over the after-effects of the visits, and stop punishing her children for no apparent reason once she gets home. As puerile as the example might sound, it is difficult to argue against the idea that somebody who knows that about herself, knows herself better than somebody who does not.

We will call these “second-order ascriptions” *nested attitudinal avowals*,<sup>3</sup> and this paper will focus on some of the features that determine the truth-conditions of a particular type of these mental-state ascriptions, self-knowledge ascriptions. In particular, we will discuss whether these nested attitudinal avowals are *authoritative*, whether the speaker’s word is all there is to say about her mental states. Authority, as the general insight that the connection between a thinker and her thoughts provides her nested attitudinal avowals with special respectability, has to be distinguished from two of its closest cousins – *transparency* and *infallibility*. Transparency, the idea that it does not make sense to ask whether a given subject A knows that *p*, where *p* is a piece of information concerning A’s mental states, is logically stronger than Authority. A theory cannot hold that self-knowledge ascriptions are both transparent and non-authoritative. If it does not make sense to ask whether A knows that she is in a certain mental state, it cannot be disputed that her sincere assertions about her mental states are always true. But Transparency and Authority are not equivalent theses. Holding Authority, a theorist can believe that every sincere assertion that A makes about her mental states is true, while still thinking that there are many things concerning A’s mental life that she might not know about, and therefore be able to deny Transparency. It makes sense to wonder whether A knows this or that about her mental life, only provided that A has never expressed an explicit opinion about it. Thus knowledge ascriptions cannot be transparent *and* non-authoritative.

Let us now turn to *infallibility*, the second claim with respect to which Authority has to be distinguished. As introduced above, authoritative nested attitudinal avowals are dangerously close to being immune to error. If these avowals are such that sincerity on the part of the speaker guarantees their truth, how can a speaker be wrong about the mental states that she says she is having? Whenever I sincerely claim to be experiencing a particular mental state, I have to be in that particular mental state, provided that my self-attributions are authoritative, in the sense described above. At times, defenders of the special status of self-knowledge have come very close to asserting some sort of incorrigibility:

One is thinking that *p* in the very event of thinking knowledgeably that one is thinking it. It is thought and thought about in the same mental act [Burge (1988), p. 70].

No errors at all are possible in strict cogito judgments; they are self-verifying [Burge (1988), p. 74].

There is a clear sense in which . . . there can be no exceptions to the claim that if someone believes that he believes that  $p$ , then he believes that  $p$ , and vice-versa [Bilgrami (1998), p. 211].

Only Burge, though, has been openly associated with this strong version of authority (see, for example, Medina (2006), p. 580), according to which our “second-order, self-ascriptive powers” [Burge (1988), p. 72] guarantee the truth of our sincere nested attitudinal avowals, or “cogito thoughts” [Burge (2013), p. 72]. We will name this kind of authority, ‘Strong Authority’:

**STRONG AUTHORITY:** knowledge self-ascriptions are said to be authoritative if sincerity on the part of the speaker and the truth of the ascription are indissociable.

But this kind of authority, according to which a sincere speaker is always a truth-teller as far as her knowledge about her own mental states is concerned, usually turns out to be a pill too hard to swallow for most theoreticians. For one, if nested attitudinal avowals were strongly authoritative, self-deception should be considered as no more than an illusion. Even Ryle presented the accommodation of self-deception as a small sacrifice for the Cartesian [Ryle (1949), p. 6], the paradigmatic defender of the authoritative character of these avowals. Second, and more crucially, these avowals are not immune to some of the problems posed by semantic externalism. The classic externalist challenge for self-knowledge is, in a nutshell, this: if the content of  $p$  is not determined internally, the truth of ‘I know that I believe that  $p$ ’ cannot be determined a priori. This is the kind of problem discussed, for example, in [Burge (1988), Bilgrami (1992b), or Nuccetelli (2003)], but an even more damning challenge to the strongest form of authority could be constructed along the same lines. Since it is not only the content of  $p$  that is externally determined, but also the content of ‘believe’, a sincere utterance of ‘I know that I believe that  $p$ ’ is not necessarily true. A speaker might think that she *believes* that  $p$ , when she in fact only *hopes* that  $p$ , *expects* that  $p$ , and such like.

Our target in this paper will thus include a weaker, and, to some, much more plausible, version of authority: *Presumptive Authority*. Typically, supporters of one kind or other of what have been called ‘transcendental arguments’ for self-knowledge [Davidson (1984); Bilgrami (1992); (1998; Moran (2001), cfr. Hacker (1997), p. 290; cfr. Brueckner (2003), p. 188, but also Wright (1998)], have defended the idea that even though our attitudinal avowals are not always authoritative, and therefore we are not infallible or incorrigible, speakers cannot be subtracted of their unique rights over their own first-person mental state ascriptions.

Despite the possibility of error, a person never loses his special claim to be right about his own attitudes, *even when his claim is challenged or overturned* [Davidson (1984), p. 104, emphasis added].

Intentional states are the kind of states which, *under ideal conditions* [emphasis added], if the possessor of a particular such state judges that he has it, then he has it. It is not as if, under these conditions, he is particularly well placed epistemologically to “track” the independently constituted facts of intentionality. No, under these conditions, his judgments constitute the facts of intentionality. [Bilgrami (1992), p. 248].

[...] positive-presumptiveness ensures that, *in all circumstances in which one has no countervailing evidence* [emphasis added], one is a priori justified in holding that the ‘no self-deceptions’ condition is justified [Wright (1998), p. 237].

These authors appear to share the assumption that, unless it is overturned by proper evidence, there is a presumption of authority regarding nested attitudinal avowals. This weaker version of authority, presumptive authority, will also be questioned in this paper.

PRESUMPTIVE AUTHORITY: even though it can be overturned by proper evidence, every time a nested attitudinal avowal is uttered, an assumption is in place that sincerity on the part of the speaker guarantees the truth of the avowal.

Davidson (1984), Bilgrami (1992) and Burge (1988) thus fall within the category of those who think that these avowals are *always* authoritative (in a strong or a presumptive sense). Others, like Medina [Medina (2006)], deny that there is anything like what Burge calls our “second-order, self-ascriptive powers” [Burge (1988), p. 72], and they maintain that the authority associated with this kind of avowals is in no way different from the kind of ‘presumption of minimal epistemic authority’ that normal speakers are endowed with in respect of their knowledge of the outside world [Medina (2006), p. 599]. We will be arguing for a third option in this paper, different from those above. Through a number of cases we will show that our intuitions concerning strong authority and presumptive authority vary widely from context to context. Sometimes knowledge self-attributions are taken to be authoritative in a strong sense, other cases only preserve the presumption of authority, and, finally, other cases will appear to contain no presumption of authority whatsoever. It is this diverse fauna that the theories we will contrast need to accommodate. The best theory would be the one that gives a simpler account of the complex phenomenon of self-knowledge ascriptions.

## II. A BRIEF PRESENTATION OF THE THEORIES IN CONTENTION

### *Truth-conditional descriptivism*

A descriptivist is committed to the idea that what makes a self-knowledge ascription true is a fact, a state of affairs that might be happening inside or outside our heads; something describable in terms of objects and first-order predicates. The facts that make our mental-state ascriptions true are analogous to those that we pay attention to when we describe our physical surroundings. My saying that I am thinking that it might rain today is true just in case I am in that particular mental state, and therefore a relation can be truly predicated between me and the proposition that it might rain today. The predicate refers to an intentional mental state, a particular relation, pointed at a proposition, as the predicate ‘being under’ refers to a particular relation established between my feet and the table in ‘My feet are under the table’.

If the descriptivist also defends that the facts that support the truth of self-knowledge ascriptions are *inner facts*, happenings in our inner stream of consciousness, then he is what Finkelstein calls a ‘detectivist’ [Finkelstein (2011), p. 5]. Non-detectivist descriptivists will then believe that the truth of self-knowledge attributions is supported by external facts, social or otherwise. Some *hybrid* positions, like Bar-On’s “neo-expressivism” [Bar-On (2004), (2009)] and, to a certain extent, Chrisman’s epistemic expressivism [Chrisman (2006), (2012); Carter & Chrisman (2012)] are designed with the purpose of keeping some the benefits of descriptivism and expressivism. Self-knowledge ascriptions serve both to express a state of mind and a factual proposition. According to “neo-expressivism”, self-knowledge ascriptions are peculiar because with them the speaker expresses a state of mind that only she is in a position to express, but they also express a proposition, whose truth conditions determine that the speaker is in the state of mind that she claims to be in. According to these views, the truth of these ascriptions hinges, at least partially, on some factual matters. So, if our interest is on the truth-conditions of nested attitudinal avowals of this kind, these positions should be placed on the descriptivist side despite their names.

### *Truth-conditional expressivism*<sup>4</sup>

Expressivism comprises a negative and a positive thesis. On the negative side, expressivism focuses on exposing the disanalogies between run-of-the-mill descriptions of physical events and mental-state ascriptions. Most psychological verbs, including ‘to know’, are sometimes used as natural-language counterparts of higher-order operators. They are functions that may take whole propositions as arguments, rather than mere objects. The verbs that we typically use to describe our surroundings correspond to first-order properties and relations, and they always take sequences of objects as their

arguments. This logical (and grammatical) difference between attitudinal avowals and descriptions plays a crucial role for the expressivist.

On the positive side, the expressivist demands that we focus on questions such as: what kind of things do we do when we ascribe mental states? When do we use these ascriptions? What kind of commitments do we acquire by using them? More specifically, here is how we propose that the expressivist insight should be positively applied to the analysis of the truth-conditions of self-knowledge attributions, and nested attitudinal avowals in general:

THE EXPRESSIVIST STRATEGY. A nested attitudinal avowal of the form ‘ $x$  knows that  $x$  Qs that  $p$ ’, where  $x$  is a thinker, Q is a psychological verb –such as ‘to think’, ‘to know’, ‘to remember’, and similar.–, and  $p$  is a proposition, would be true, when uttered in a suitable context  $c$ , if and only if a contextually salient set of features only makes sense if the avowal is not taken to be false.

Sometimes the most contextually salient feature will be the speaker’s avowal, some other times the importance of the speaker’s ascription would be overrun by other features of the context, such as the physical circumstances surrounding the ascription, other propositional attitudes attributed to the speaker, her reputation, and such like. We will see the strategy at work in the last section of the paper.

### III. AUTHORITY AND TRANSPARENCY OF SELF-KNOWLEDGE ASCRIPTIONS: A CASE-BY-CASE STUDY

#### Case 1: *Strongly Authoritative vs Non-Authoritative Intuitions*

In an episode of the classic TV series *The Practice*, the main character of the series, the lawyer Robert Donnell, faces a hearing to determine his responsibility in the deaths of several police officers. Donnell is romantically involved with a district attorney. While at her house, Donnell unintentionally overhears a phone call made by the police to the district attorney, informing her of a future raid on the premises of a well-known gangster, Mr Eisikov. The gangster happens to be one of Donnell’s clients. Once the lawyer gets the information, he feels that his duty is to inform his client, no matter how fortuitously he has run into this piece of knowledge. He tips off his client, and when the police arrive at his warehouse, they are ambushed; some of them are killed. A hearing is called by a judge to determine whether Donnell is responsible for the deaths of those police officers. The decision ultimately hinges on Donnell’s state of mind when he passed on the information. He

will be charged if it is safe to assume that he knew what was likely to happen once he told the gangster about the raid. Otherwise, he will walk free. Here is the exchange, during that hearing, between the prosecutor, the district attorney acting here as a witness, and the defense attorney.

PROSECUTION: You have a personal relationship with the defendant?

WITNESS (district attorney): We date. That's why he was in my apartment when the phone call came in. That's how he overheard we were scheduling a raid on Mr Eisikov's safe house.

PROSECUTION: So the defendant knew about the raid.

WITNESS: I'd answer, but the defense is about to object.

DEFENSE: Objection, the witness is incompetent to say what the defendant may or may not have known.

PROSECUTION: After the shootout, did you and the defendant have a conversation?

WITNESS: Yes.

PROSECUTION: And did he tell you whether he informed Mr Eisikov in advance that a police raid was coming?

WITNESS: He told me that he felt legally obligated to reveal the information to his client. Yes.

PROSECUTION: Thank you.'

While the prosecutor tries to establish that Donnell knew what was likely to happen by questioning his lover, both the witness and the defense attorney make it abundantly clear that only the defendant, Donnell, is in an appropriate position to say what he did or did not know at the moment of the crime ('the witness is incompetent to say what the defendant may or may not have known'). Only he can ever say what was going through his mind at the time, nobody else's testimony should be taken into consideration on this matter. Unhappy with this response, the prosecutor continues pressing the issue. Those involved in the situation, he argues, those who are aware of the facts of the case, must have an opinion as to whether the defendant knew what was likely to occur once the gangster received the tip-off. These opinions are not mere raw subjective intuitions on the issue, they are reasonable, informed, perspectives on the situation. They are the results of flawless processes of reasoning undertaken by knowledgeable people trying to make sense of the facts. Donnell knew that his client was a dangerous man, and knew that he was likely to act on the information he received, no matter what he now says he knew at the material time. Presumably, only knowing that his client would act as the result of receiving the information made it possible for Donnell to feel that it was his duty to tell his client. And how, the prosecutor seems to reason, would a person know that a gangster would act on the information that there was going to be a police raid on one of his premises, without knowing that the result of this action was likely to involve the deaths of police officers? The defendant *must have known* what was likely to happen.



A contrast can be appreciated here between the defense attorney, who thinks that only what the defendant knew *for a fact* can be counted on as evidence, and the prosecutor, who argues that clearly determining what the defendant *must have known* is evidence enough to charge him with the crime. The defense attorney, playing the role of the descriptivist, maintains that only correctly establishing the facts can lead to a conviction –in this case, facts about the mental life of the defendant to which only he has access, while the prosecutor claims that nothing but what is necessary to make sense of the situation is needed in order to proceed with the charges. The prosecutor’s line of thought does not conclude with a factual statement, but with an appreciation regarding what *has to be* the case for the whole situation to make sense.

The judge not only faces a decision on Donnell’s fate as a lawyer, but she also suddenly has the opportunity to take a stand in a philosophical debate! As we see from her words, the ruling did not appear to be easy:

I called you back in because I made up my mind, but in the two minutes that it took you to file back in I changed it. Whether or not you’re guilty here really depends on exactly where you were in your head when you talked to your client, whether you knew what would happen, or what you thought might. Unfortunately for me as a judge, that answer still lies within your head, and your head alone. On the facts: Dismissed without prejudice, the defendant is free to go.

So, finally, the judge takes sides with the defense attorney, and Donnell is released. Only the facts can be used to decide on legal matters, the judge argues, and she is therefore forced to set the defendant free, provided that only the defendant is considered to have access to those facts, occurring as they might have had ‘within his head’. She nevertheless seems to have given some thought to the idea that in certain cases making sense of a set of facts happening ‘outside the head’ can be enough to determine what *must have been* ‘inside it’, and rule accordingly.

Both the defense attorney and the judge give the upper hand to the detectivist in this case. Only the subject can know what he was thinking at some point or another. The truth of a knowledge ascription (‘I know that I did not know back then what was likely to happen’<sup>5</sup>) is only granted by an internal fact, which can solely be accessed through internal epistemological channels. Only Donnell’s sincere confession could secure his conviction, since only he has access to whatever was going through his head at the time. The judge and the defense attorney adopt a factulist approach to the truth of knowledge ascriptions, *and* give them the traditional status associated with Strong Authority. They are willing to reject as relevant any amount of external information that the prosecutor adduces concerning the circumstances of the case, and thus take the speaker not only to have a presumption of authority, but also to be strongly authoritative. The prosecutor, on the other hand, is not trying to appeal to any sort of external fact – or any fact, for that matter –, neither

strong authority nor presumptive authority seem to be an issue for him. His point is simply that, given the circumstances, and no matter what he might end up saying, the defendant *must have known* what was likely to happen, and should therefore be held responsible for the deaths of those police officers. Donnell's word is just not the most relevant feature in order to determine whether he knew or not. Even though the judge rules against him, the prosecutor's non-descriptive assessment of the situation needs to be accounted for, as are the doubts that the judge expresses in her ruling.

The first case might have left the reader with mixed intuitions concerning which should be the appropriate approach to the analysis of knowledge self-ascriptions. As it happens, it is presented in the series as a source of moral discomfort, as a means to provoke some reflection on the nature of legal obligation. In the following case, our intuitions should be clearer.

*Case 2: Presumptive Authority, but Not Strong Authority.*

In one of the episodes of Werner Herzog's *On Death Row*, the crimes of James Barnes, sentenced to death, come under scrutiny. James Barnes was arrested and accused of murdering his wife. Under interrogation, he presented the event as an accidental death. They were fighting, he said, and she was really nervous; while trying to calm her down he grabbed her by her neck, and they fell together while he was still holding her. Only afterwards did he realize that she was dead. He claimed to have no intention of murdering her. In spite of Barnes' testimony, the detective in charge of the investigation, Tod Goodyear, considers that discarding the possibility of accidental death and working on the hypothesis of a murder is the only way to make sense of an overwhelming amount of features of the circumstances surrounding the death of the woman, and therefore concludes that James Barnes *must have* intentionally killed his wife. The jury agrees with him, and James Barnes is convicted and given a life sentence. In the detective's own words:

He tried to play it off in the interview that it was all an accident. That he had gotten mad at her, he had put his choke hold on her to calm her down, because she was getting out of control, and in doing so they fell to the floor, and when they fell to the floor he said he either must have broken her neck or in some form choked her. Now as we go through this interview and keep going on, he tells us and describes it quite vividly how he picked her up off the ground and to the point where she's shaking and he notices that ... [crude details left out], which is a little bit different than 'we fell to the ground and this was an accident.' So as the interview goes on he gets much more descriptive, and details you a very violent act, a very vivid violent act, in his memory of noting that she's shaking, of noting that [crude details left out], so something that he was cognizant of what was going on at that point in time. And it was obvious in my mind, and of course it was obvious to the legal system, that this was a premeditated act. At that point he was there to kill her.

Tod Goodyear is not dismissing Barnes' testimony right away. We do not even need to suppose that the detective is not taking Barnes to be a man of his word; the testimony of the suspect is simply not enough this time to settle the case and grant the innocence of the suspect. There are so many details in Barnes' story that do not seem to fit the original self-ascription, that an alternative hypothesis, one in which Barnes *must have known* what he was doing, finds its way into the courtroom successfully.

The detective in Barnes' case, as the aforementioned prosecutor, adopts a strategy very close to what we have described above as the expressivist approach to the meaning of attitudinal avowals. Barnes' testimony, even if sincere, does not *per se* guarantee his innocence, and thus his testimony is not taken to be authoritative. But his words have to be counterbalanced by overwhelming contrary information, overturned by proper evidence, and therefore a presumption of authority is still in place. A conviction can only be obtained because the jury and the judge agree with the detective in that, all in all, the defendant's word cannot be all there is to say with regard to his own intentions. Here nobody plays the role of the descriptivist to defend Barnes.

It could perhaps be argued here that the detective is not merely second-guessing the alleged murderer's statement, wondering about the truth of the ascription, but also assuming that he is lying. Once he has sufficient evidence, he concludes that Barnes has been lying in his description of the event as an accident, and that is how he reaches his conclusion that Barnes is in fact a cold-blooded killer. We think that Goodyear's words can be perfectly understood as casting some doubts on Strong Authority, rather than on the man's sincerity, and that any extra assumptions concerning Barnes' sincerity might be an over-interpretation. For those willing to stick to the hypothesis that Goodyear simply took Barnes to be a liar, here is an analogous case in which it is precisely the accused's pristine reputation that motivates the audience to question Strong Authority.

In Dostoyevsky's *The Brothers Karamazov*, the Elder Zosima tells the story of Mikhail, a highly respected member of his community, married with two children, who, inspired by Zosima's conversion, decides to publicly confess to a crime that he committed 14 years earlier. Feeling betrayed by the woman he was in love with, Mikhail breaks into her house one night, kills her in her sleep, and steals some letters and a few worthless items to make it look like a common robbery. Due to some fortunate turn of events, the woman's servant is accused of the crime, and no one ever suspects Mikhail. He carries the guilt for many years, until he decides to make a public confession, which Zosima encourages. Surprisingly, the city is unconvinced by Mikhail's confession, they even seem unmoved by the physical evidence of the crime that he has kept for 14 years, and now produces in front of everybody.

He had taken both letters... but why? Why had he kept them for fourteen years instead of destroying them as evidence? And what happened then: everyone was astonished and horrified, and no one wanted to believe it, though they listened with great curiosity, but as to a sick man, and a few days later it was all quite decided among them, the verdict being that the unfortunate man had gone mad. The authorities and the court could not avoid starting proceedings, but they also held back: though the articles and letters he produced did make them think, here, too, it was decided that even if the documents proved to be authentic, a final accusation could not be pronounced on the basis of these documents alone. And the articles he might have obtained from the woman herself, as her acquaintance and trustee [*The Borthers Karamazov*, 1992 edition, translated by Richard Pevear and Larissa Volokhonsky, p. 264].

Note that Mikhail's audience not only refuses to believe that he committed the crime, but they decide to take him for a sick man as well. Not only do they believe it impossible that he committed such a horrible crime, but also that he would be capable of fabricating such a story, given his prior status and respectability. They decide that, even though he cannot be taken for a liar, his words cannot constitute truth, and thus he must be deranged. Mikhail must be mistaking his own memories for some unnerving fantasy, a product of an ill mind. Mikhail's statements are not to be trusted any longer, since he is unable to think clearly about his own memories; he is confused about the status of some of his mental states.<sup>6</sup> Authority is, therefore, left aside by the community in order to safeguard a wide array of assumptions concerning the man, his reputation, the nature of the crime, the moral health of the community, and so on. The community starts from a presumption of authority, and overturns it once it has taken into consideration enough features of the case. Obviously, the community is wrong in not taking Mikhail at his word since he claims to be remembering a crime that he actually committed, but are they unreasonable in so doing? The strategy undertaken by the authorities in Mikhail's case is not that different from the one of Detective Goodyear. They decide to pay attention to the complex set of details of the circumstances surrounding the case, and try to make sense of them. Sometimes this strategy will take you closer to the truth, and other times, very occasionally, it can point you in the wrong direction, but there is no way to know a priori that preserving Strong Authority against all odds is going to guarantee the safest epistemological results.

In case 1, the defense and the prosecution exhibited different intuitions with respect to the status of the theses which we are discussing in this paper. The defense wanted to preserve Strong Authority to exonerate their client, while the prosecutor did not even bother arguing against a presumption of authority, several other features of the situation were decisive for him. In case 2, on the other hand, both the detective and the community undertook the task of overturning authority, starting thus from a presumption of authority, in order to demonstrate

that the truth of a certain statement is not guaranteed by the speaker's sincerity. Let us now finally turn to a case in which Strong Authority is preserved.

Case 3: *Strongly Authoritative Self-Knowledge Ascription*

On July 24<sup>th</sup> 2013, in Galicia, Spain, one of the worst train accidents in Spanish history took place. The train, which should have been running at 80 km/h, took a dangerous turn at almost 200 km/h. It crashed and 79 people died. The police and the judge tried to determine whether the driver, Francisco José Garzón, was responsible for the crash, and it seemed to be of vital importance to determine what the mental state of the driver was at the time of the accident. Mechanical failure could only be ruled out, and the driver charged, if it were established that the driver had not been paying proper attention to the controls. The police and the public prosecutor, Antonio Roma, gathered as much evidence as possible from the scene, and the driver was interviewed:

PROSECUTOR: What were you thinking right before entering the second tunnel?

DRIVER: I don't know, if only I knew... the burden I'm going to be carrying for the rest of my life is enormous.

PROSECUTOR: Everybody here, the police and the rest of us, is working to determine what went through your head. I beg you to make an effort...

DRIVER: Your Honor, I sincerely tell you that I don't know. I'm not crazy enough not to pull the breaks.

(This dialogue is a literal translation of the one appearing in the piece "El maquinista del tren de Santiago: 'No lo vi. No estoy tan loco para no frenar'", by José Precedo, published in the Spanish newspaper *El País*, July 31st 2013.)

Without reason to suspect he was lying, the prosecutor tried to obtain a definitive confession or an exonerative declaration from the driver, who continued to claim that he did not know what was going through his mind at the time. Why did he not stop the train? What was he thinking? Was he distracted? The answer to these questions depended on the state of mind that the driver was in at the time of the accident, but the driver insisted that he did not know. The prosecutor took a claim equivalent to 'I do not know whether I knew at the time what the speed was' to be meaningful, and thus *rejected transparency*. He tried to obtain an answer from the driver, nevertheless, because he thought that Garzón was in the best possible position to know what was going through his mind at the time –Roma believed the self-knowledge ascription to be *strongly authoritative*.

IV. CONCLUSION. EXPRESSIVISM, DESCRIPTIVISM AND AN ARGUMENT FROM PARSIMONY

Let us take stock. As we saw in the first section, philosophical intuitions with respect to Strong Authority and Presumptive Authority can be categorized

into three groups, depending whether the theorist believes that self-knowledge ascriptions are (i) always authoritative (in one way or another), (ii) never authoritative, or (iii) sometimes strongly authoritative, sometimes only linked to a presumption of authority, and sometimes non-authoritative. We take the previous section to have provided evidence in favor of the third option. In case 1, while the judge and the defense attorney thought that only the defendant could truly know what he was thinking – and therefore took his putative self-ascription to be strongly authoritative –, the prosecutor rejects that, not even feeling compelled to argue against a certain presumption of authority. In the series, the case allegedly poses a genuine conflict, one in which none of the parties should be taken to be patently right or wrong. In case 2, Detective Goodyear and Mikhail's community take the self-ascription under discussion to be non-authoritative, in a strong sense. Barnes and Mikhail's recollection of the events, expressed through self-ascriptions of their memories, are not taken at face value. It is doubted that those are genuine memories, and so the nested attitudinal avowal 'I know that I remember that...' is not true simply because the subject who is supposed to experience those mental states expressed it. But Detective Goodyear and Mikhail's community need to gather evidence in order to overturn authority, and so we can safely assume that they start from a presumption of authority. Finally, the words of the driver of the train in case 3, during his interview, are taken to be strongly authoritative. The purpose of this section is to examine the theoretical resources that descriptivism and expressivism have in order to deal with the intuitions behind the cases presented above.

Descriptivism, as introduced above, involves a commitment to the idea that the truth of self-knowledge ascriptions has to be supported by a fact, internal or external. Detectivist and non-detectivist descriptivist seem to be associated with options (i) and (ii). Either the fact we are seeking is internal, as the detectivist maintains, and thus self-knowledge ascriptions are always authoritative, or the fact sought is external, as the non-detectivist descriptivist claims it is, and self-knowledge ascriptions are not authoritative. As it stands, the descriptivist seems unable to provide a suitable explanation for the intuitions behind cases 1, 2, and 3, which arguably support option (iii), namely that sometimes self-knowledge ascriptions are authoritative, and sometimes they are not.

Expressivism, on the other hand, is able to apply its strategy homogeneously through the cases. The genuine debate between the defense attorney and the prosecutor in case 1 could be explained as a difference in determining the salient features of the context that the truth of self-knowledge ascriptions has to help make sense of. While the defense attorney thinks that the most salient feature of the context is the fact that the defendant claims not to have known at the time about the lethal consequences of his action, the prosecutor insists on the idea that there are many other features of the circumstance that do not make sense if the self-ascription is taken to be true. The detective and

the authorities in Barnes' case consider that the suspect's vivid recollection of the violent act, together with some other more general assessments regarding the character of the suspect, are more salient than his exculpatory self-ascription. For Mikhail's community, his spotless reputation and the kind of moral values associated with it are too high a price to pay to safeguard the truth of the self-knowledge attribution, so they are contextually more salient, and making sense of them turns out to be the priority. While reading *The Brothers Karamazov*, we learn many other details of the crime, and we know that the community made the wrong estimation of the salient features of the context, but it is only with the aid of a sufficiently flexible theory, such as expressivism, that we can explain why the judgment of the community was mistaken, and why it was not entirely unreasonable for them to have made such a judgment.

So far, we have only considered *strict* versions of descriptivism. Should we not allow the descriptivist theories some degree of *contextual flexibility*? Would *liberal* descriptivism be in a better position to deal with cases 1, 2 and 3? Detectivism, non-detectivist descriptivism, and "neo-expressivism" could in principle be sensitive to changes in context, so that sometimes the fact that they are seeking would be internal, and some other times it would be external. With this dose of context-sensitivity, liberal detectivism and liberal non-detectivist descriptivism would be equivalent theses. A detailed exploration of the pros and cons of this theoretical move would greatly exceed the scope of this essay, but this does not prevent us from suggesting a final argument in favor of expressivism. Any liberal-descriptivist theoretical option is going to have to provide (1) theoretical resources to sharply distinguish between contexts in which the relevant fact is to be found *in the inside*, and contexts in which we have to look for the fact that supports the truth of the ascription *outside*, and (2) accordingly, two different semantic strategies to give the truth-conditions of self-knowledge ascriptions. Only then can a liberal descriptivist reach the explanatory power of the expressivist. The expressivist, on the other hand, manages to account for the variety of intuitions involved in the cases we have described with a single theoretical stroke. The unitary semantic strategy of the expressivist makes his or her theory more parsimonious. It is, therefore, a better theory.

*Departamento de Filosofía I  
Universidad de Granada  
Campus de Cartuja  
18071, Granada, Spain  
E-mail: nef@ugr.es*

## ACKNOWLEDGMENTS

The author would like to thank Manuel de Pinedo, María José Frápolli, Manuel Heras, Andrés Soria, Jesús Vega, David Finkelstein, and two anonymous referees of *Teorema* for their useful comments. Research leading to this paper has been supported by the project FFI2010-15704, funded by the Ministerio de Ciencia, Educación e Innovación de España, and the project *Expresivismo Doxástico*, funded by the Universidad de Granada.

## NOTES

<sup>1</sup> This use of ‘transparency’, which can be found in [Wright (1998), p. 15], should be distinguished from the notion of ‘transparency-to-the-world’ that Bar-On takes from Evans (see Bar-On (2004), p. 105).

<sup>2</sup> Even though the characterization of these features may slightly vary in the literature, we take the ones that we include here to be neutral enough, so that nobody finds it problematic to follow our arguments due to terminological issues.

<sup>3</sup> There is a reason why we do not follow the usual terminology here. Every propositional attitude verb, such as ‘to believe’, ‘to know’, and so forth, performs the function of a second-order operator in natural language, given that they take as arguments complete propositions, rather than mere objects. In this sense ‘second-order ascriptions’ is not descriptive enough for the linguistic phenomenon that we try to grasp. *Nested attitudinal avowals*, on the other hand, captures the idea that we have two psychological verbs introducing two propositional attitudes, and one of them works as the argument of the other.

<sup>4</sup> It might seem somewhat odd to present expressivism as a positive thesis with respect to the truth-conditional analysis of certain expressions. Classic expressivism is commonly associated with the idea that the semantics of the expressions on which it shines the spotlight—ethical statements, normative statements in general, epistemic statements, etc. — are no different from that of grunts and cries. These expressions do not express propositions, so they cannot be declared true or false, they are not “truth-apt”. This assumption is not only historically dubious but also a non sequitur. From the claim that certain expressions do not modify the truth-conditions of the expressions under their scope it does not follow that the whole expression lacks truth-conditions [Cfr. Frápolli & Villanueva (2012)]. At some point, considering the reactions that it usually elicits, one is tempted to stop using the label ‘expressivism’ for this position, and to invent a new, perhaps more neutral, name for it, but this would be betraying its historical origins, from Ayer to Gibbard, in a way which seems to be completely unjustified.

<sup>5</sup> Note that, unlike the vast majority of cases discussed in the literature regarding authority, most of the examples contained in this paper are not expressed in the present tense. This, however, does not introduce any extraneous element into the discussion. Our discussion is not about memory. What is debated in each case is not whether the speaker remembers correctly what he or she thought at a given point in time, but whether he or she is the sole and fundamental source of authority with respect to what was going through his or her mind at the material time. Unfortunately



for the philosopher, real life does not provide many examples of nested attitudinal avowals used in the present tense.

<sup>6</sup> Again, here the problem does not concern how good his memory is, but whether a statement of the form ‘I know that I Q’s p’ (I know that I remember p) can be taken to be true simply because the man is sincere. The community does not believe that Mikhail lacks a good memory, but that he is sick. Only in extreme cases is a bad memory considered a mental illness.

## REFERENCES

- BAR-ON, D. (2004), *Speaking My Mind*; NY, Oxford University Press.
- (2009), “First-Person Authority: Dualism, Constitutivism, and Neo-Expressivism”, *Erkenntnis*, 71, pp. 53-71.
- BILGRAMI, A. (1992), *Belief and Meaning*; Oxford, Blackwell.
- (1992b), “Can Externalism be Reconciled with Self-Knowledge?”, *Philosophical Topics* 20 (1), pp. 233-267.
- (1998), “Self-Knowledge and Resentment”; in: Wright et alii (eds.) (1998), 207-243.
- BRUECKNER, A. (2003), “Two Transcendental Arguments Concerning Self-Knowledge”; in Nuccetelli (ed.) (2003), pp. 185-200.
- BURGE, T. (1988), “Individualism and Self-knowledge”, in: *Journal of Philosophy* 85, pp. 649-663.
- (2013), *Cognition through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection: Philosophical Essays*, Volume 3, Oxford, Oxford University Press.
- CARTER, J. A. and CHRISMAN, M. (2012), “Is Epistemic Expressivism Incompatible With Inquiry?”, *Philosophical Studies*, 159, pp. 323-339.
- CHRISMAN, M. (2007), “From Epistemic Contextualism to Epistemic Expressivism”, *Philosophical Studies* 135, 225-254.
- (2012), “Epistemic Expressivism”, *Philosophy Compass* 7 (2), pp.118-126
- DAVIDSON, D. (1984), “First Person Authority”, *Dialectica* 38, pp. 101-111.
- FINKELSTEIN, D. (2003), *Expression and the Inner*, Cambridge, Mass: Harvard University Press.
- FRAPOLLI, M., and VILLANUEVA, N. (2012), “Minimal Expressivism”, *Dialectica* 66, 4, pp. 471-487.
- HACKER, P. M. S. (1997), “Davidson on First-Person Authority”, *The Philosophical Quarterly*, 47, 188, pp. 285-304
- MEDINA, J. (2006), “What’s so Special about Self-Knowledge”, *Philosophical Studies*, 129, 3, pp. 575-603.
- MORAN, R. (1997), “Self-Knowledge: Discovery, Resolution, and Undoing.” *European Journal of Philosophy* 5, pp. 141-163.
- . (2001), *Authority and Estrangement*, Princeton: Princeton University Press.
- NUCCETELLI, S. (ed.) (2003), *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, Mass., MIT Press.
- RYLE, G. (1949), *The Concept of Mind*, Chicago, The University of Chicago Press.
- WRIGHT, C. 1998. “Self-Knowledge: The Wittgensteinian Legacy”, in Wright et alii (eds.) (1998), pp. 101-122.
- WRIGHT, C, SMITH B., and MACDONALD, C. (eds) (1998), *Knowing Our Own Minds*, Oxford, Oxford University Press.