INNER SPEECH AND METACOGNITION: A DEFENSE OF THE COMMITMENT-BASED APPROACH

Víctor FERNÁNDEZ CASTRO

ABSTRACT: A widespread view in philosophy claims that inner speech is closely tied to human metacognitive capacities. This so-called format view of inner speech considers that talking to oneself allows humans to gain access to their own mental states by forming metarepresentation states through the rehearsal of inner utterances (section 2). The aim of this paper is to present two problems to this view (section 3) and offer an alternative view to the connection between inner speech and metacognition (section 4). According to this alternative, inner speech (meta)cognitive functions derivate from the set of commitments we mobilize in our communicative exchanges. After presenting this commitment-based approach, I address two possible objections (section 5).

KEYWORDS: inner speech, metacognition, commitments

1. Introduction: Talking to Oneself

Metacognition or thinking about thinking is a fundamental human cognitive capacity.[1] This capacity is devoted to evaluating, predicting or modifying our cognitive performances, so it endows us with a unique cognitive and behavioral flexibility and adaptability. Several authors have claimed that there is a constitutive connection between these metacognitive capacities and the linguistic ability of talking to oneself,[2] so humans are able to flexibly modify, regulate and access their cognitive processes because they are able to structure their own mental states in a linguistic format through self-directed talk. This so-called *format view of inner speech*[3] claims that capturing our mental states in linguistic format allows

[1] Michael T. Cox, Anita Raja, and Eric Horvitz, eds., *Metareasoning. Thinking about Thinking* (Cambridge, Mass.: MIT Press, 2011); John Dunlosky and Janet Metcalfe, eds. *Metacognition* (Los Angeles: SAGE Publications, 2019).

[2] Daniel C. Dennett, *Consciousness Explained* (London: The Penguin Press, 1991); Ray Jackendoff, *The architecture of the language faculty* (Cambridge, Mass.: MIT Press, 1997); Andy Clark, *Being there* (Cambridge, Mass.: MIT Press, 1997); Jose Luis Bermudez, *Thinking without words* (Oxford: Oxford University Press, 2003).

[3] Fernando Martínez-Manrique and Agustín Vicente, "The activity view of inner speech," *Frontiers in psychology* 6 (2015): 232.

us to acquire the metarepresentational capacities underlying the unique human metacognitive competence of modifying and accessing our own mental states

The aim of this paper is to defend an alternative to the format view as a theory of the connection between inner speech and metacognition. The alternative I put forward is based on a Commitment-Based approach to communication and inner speech according to which the main purpose of communication is to establish commitments and entitlements to coordinate agents; so, the cognitive function of inner speech derivate from this social function of outer speech. The structure of the papers goes as follows: Firstly, I present the format view along with two objections (section 2 and 3). These objections challenge two central ideas of the format view: (1) the notion of metacognition as access, and (2) the idea that metacognition requires metarepresentations. In section 4 and 5, I introduce the commitment-based view and how it can account for the different cognitive functions associated with metacognition. Finally, in section 6, I address two possible objections to the alternative.

2. The Format View of Inner Speech

Inner speech is often defined as the phenomenon we experience when talking silently to ourselves. The contemporary interest on the phenomenon starts with the publication in English of the work of the Soviet psychologist Lev Vygotsky who, after realizing that children systematically talk to themselves out loud (private speech), started to study the role of private and inner speech in the development of high cognitive capacities.[4] In contemporary psychology, the research on private and inner speech has resulted into different studies that connect inner speech with different cognitive capacities, including conscious control, working memory and attention.[5]

Besides this empirical evidence, there are different debates on the format, nature, and function of inner speech.[6] The fundamental question underlying those

[4] Lev S. Vygotsky, *Thought and language* (Cambridge, Mass.: MIT Press, 1984, Original work published 1934).

[5] Rafale Diaz and Laura Berk, eds. *Private speech: From social interaction to self-regulation* (Hillsdale, N.J.: L. Erlbaum, 1992); Daniel Gregory "Inner speech, imagined speech, and auditory verbal hallucinations,"*Review of Philosophy and Psychology* 7,3 (2016): 653–673; Adam Winsler, Charles Fernyhough and Ignacio Montero, eds. P*rivate speech, executive functioning, and the development of verbal self-regulation* (Cambridge: Cambridge University Press, 2009).

[6] Martínez-Manrique and Vicente, "What the...! The role of inner speech in conscious thought,"*Journal of Consciousness Studies* 17 (2010): 141–167; Keith Frankish,"Evolving the linguistic mind,"*Linguistic and Philosophical Investigation* 9 (2010): 206–214;Peter Langland-Hassan, "Inner speech and metacognition: in search of a connection,"*Mind & Language*29 (2014):

debates is why do we talk to ourselves? A widespread answer in philosophy of mind maintains that we talk to ourselves in order to display metacognitive abilities, that is, we talk to ourselves to consciously access our own thoughts.[7] This so-called format view of inner speech associates the function of our self-talk to some structural features of the linguistic format. The main thesis is that language, in virtue of these features, is the only representational vehicle that allows codifying mental states in a way that can be objects of further thoughts. In other words, language facilitates what Clark calls second-order dynamics. Language codifies thoughts that can be brought into working memory in a way that attention can be directed to them, and thus, be objects of conscious access. Although these authors share the perspective of inner speech as a metacognitive facilitator, they differ about which properties make language appropriate for such function. In this sense, for instance, Clark argues that the features of language that allows us to recruit it for cognitive purposes are its context-dependency and neutral modality.[8] On the other hand, Bermudez considers that, given that all conscious access must be carried out on perceptual modality, language is the only representational vehicle that allows personal level conscious access and is, at the same time, a structured vehicle. Contrary to other personal vehicles as images, language is structured and compositional. Contrary to other structured vehicles as mentalese inner speech is a vehicle we can consciously access.[9] Thus, inner speech is the only representational format that facilitates second-order dynamics to conceptually structured thoughts.

This picture on inner speech face several problems related with some of its fundamental theses.[10] However, the aim of this paper is to reveal the problems of the view regarding two fundamental assumptions; namely, how the model assigns a central role to metarepresentations in metacognitive capacities, and how metacognition is understood in terms of access to mental states or processes. First, according to the format view, when an agent experiences an episode of inner

511– 533; Peter Langland-Hassan and Agustin Vicente, eds. *Inner Speech: New Voices* (USA: Oxford University Press,2018).

[7] Jose Luis Bermudez, *Thinking without words*; Andy Clark, *Being there*; Daniel C. Dennett, *Consciousness Explained*; Jackendoff, *The architecture of the language faculty*.

[8] Clark, *Being there,* 178.

[9] Jerry Fodor, *The language of thought* (Cambridge, Mass.: Harvard university press, 1975)

[10] Apart from Martínez-Manrique and Vicente, "The activity view of inner speech," the problems of the format view has been emphasized by Marta Jorba and Agustin Vicente, "Cognitive phenomenology, access to contents, and inner speech," *Journal of Consciousness Studies* 21, 9-10 (2014): 74-99; Víctor Fernández Castro, "Inner Speech in Action," *Pragmatics & Cognition* 23, 2 (2016): 238-258; or Bart Geurts, "Making sense of self talk," *Review of Philosophy and Psychology* 9, 2 (2018): 271-285.

speech, for instance when someone utters silently a sentence such as 'the unemployment in Europe have decreased at the expense of worker's rights,' she can access her own mental state because, through the access of this internal episode, she can infer the state that she believes that the unemployment in Europe have decreased at the expense of worker's rights. So, metacognition requires forming representations about that mental state in order to perform other actions as controlling or regulating the state in question. This metacognitive capacity can be understood as a device that takes the content of an inner speech episode as an input and produce a metarepresentational state of the form 'I believe (desire, imagine) that P' as an output. Likewise, inner speech episodes allow us to access to our mental states as far as facilitates the generation of metarepresentations with the form 'S verbs P.' Understanding metacognition in metarepresentational terms is not new. As Proust has shown, considering that metacognitive capacities rely upon the ability to form metarepresentation is widely shared assumption in cognitive sciences and philosophy.[11] The innovation of the format view, then, is connecting these metarepresentational capacities to inner speech and the capacity of putting thoughts in a linguistic format.

Second, the format view is strongly committed to a particular notion of metacognition as access.[12] Again, as Proust argues, most of the philosophical approaches to metacognition in philosophy and cognitive sciences assume that the second-order regulation and control of cognitive processes require the subject to be able to access, either through introspection or inference, to the contents of the first level processes and states. So, humans could not regulate, evaluate and modify their first-order mental processes and states without having access to such processes and states. In the format view, capturing our thoughts through inner episodes allow us

[11] Joëlle Proust has examined this and other aspects the standard view of metacognition (see Joëlle Proust, "Metacognition," *Philosophy Compass* 5, 11 (2010): 989-998; *The philosophy of metacognition: Mental agency and self-awareness* (Oxford UK: Oxford University Press, 2013). She mentions as proponents of such standard view to John Flavell, "Metacognition and Cognitive Monitoring: A New Area of Cognitive- Developmental Inquiry," *American Psychologist* 34 (1979): 906–911; Alan Leslie, "Pretense and Representation: The Origins of Theory of Mind,'' *Psychological Review* 94 (1987): 412–26; Josef Perner, *Understanding the Representational Mind* (Cambridge, Mass.: MIT Press, 1991); Alison Gopnik, "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality," *Behavioral and Brain Sciences* 16, 1 (1993): 1–15; Peter Carruthers, "Meta-cognition in Animals: A Skeptical Look," *Mind and Language* 23 (2008): 58–89; "How Do We Know Our Own Minds: The Relationship between Mindreading and Metacognition." *Behavioral and Brain Sciences* 32 (2009): 121–82

[12] Joëlle Proust,"Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition?," *Synthese* 159, 2 (2007): 271-295.

to access our mental states and processes because we can self-ascribe such states by forming metarepresentation of the form 'I verb P.' So, inner speech episodes facilitate the second-order access our metacognitive capacities consist in.

3. Telepaths and the Young Rich Communist

This section brings into focus two objections of the format view, which lay on the two aforementioned assumptions. That is, the idea that metacognition must be understood in terms of access and the idea that metacognition is carried out in a metarepresentational format. These two objections prepare the ground for defending the commitment-based approach I characterize in the next section.

The first problem to the format view lies on the restricted power of the notion of metacognition as access to account for how inner speech make a difference for explaining the cognitive and behavioral flexibility associated to metacognition. In principle, the explanandum of a theory of this type must be to explain how inner speech, as long as it endows linguistic creatures with certain metacognitive capacities, can account for some of the patterns of actions and mental skills associated with thinking about thinking, for instance, cognitive flexibility or the capacity to evaluate and regulate actions. However, the format view seems to fail to achieve this objective. Although the format view gives a reasonable explanation of how a creature can access to her mental states, it is hard-pressed to explain how this access is translated into certain special cognitive abilities. For instance, why the metacognitive capacities associated with inner speech facilitate the rise of cognitive regulation or flexibility. Part of the obstacle a defender of the format view must address is that, although the position claims that inner speech brings certain mental states into consciousness, it does not explain how this 'bringing mental states into consciousness' plays a role in regulating or evaluating first- order processes. As McGeer argues, having access to our own mental states would play a role analogous to the role of a telepath that could read our mind, seeing our mental states and processes, but could not exercise any type of power to modify or regulate them.[13] If the format theory aims to explain which function the inner speech plays in the acquisition of metacognitive capacities, the theory should not only explain how certain distinctive mental states or processes are produced, but also how accessing those states and processes make a difference for the type of abilities we usually associate with metacognition (control of attention, regulation, cognitive flexibility).[14] In other words, monitoring our

[13] Victoria McGeer, "The Moral Development of First-Person Authority," *European Journal of Philosophy* 16, 1 (2007): 81-108.

[14] See Proust, *The philosophy of metacognition,* 29-78.

mental states is not sufficient for explaining the cognitive and behavioral flexibility associated with metacognition, and thus, the format view must be regarded as incomplete.

A possible way out to this problem may appeal to the notion of metarepresentation. The defender of the format view could argue that the metarepresentational states that inner speech produces could modify certain pattern of behavior and cognition in a flexible way. For instance, image a physicist on the way home to finish an article that the editors of a journal have been waiting for. At the moment, she is entering her house, an utterance crosses her mind 'the dinner!' Suddenly, she remembers she has invited some friends for dinner and the fridge is empty. 'I gotta go to the grocery store.' The physicist changes her route and stops at the grocery store before going home. According to the format view, inner speech episodes could allow the agent to access her mental states (remembering that she has organized a dinner, the belief that the fridge is empty and the belief that she must go to the grocery store) in a way that she can abort her action of going home and trigger the action of walking toward the store.

However, this solution does not solve the problem. Notice that explaining how behavioral and cognitive flexibility derivate from inner speech does not seem to necessarily rely on metarepresentational states. In principle, the physicist's cognitive processes can be carried out by first-order processes. The appropriate behavioral pattern can be triggered by bringing out the appropriate information without a self-ascription of the given mental states; for instance, bringing out the information that she should go to the store and that she has a dinner tonight rather than the self-attribution of such mental states. It is the mental states per se and not the self-attribution of these states what seems to play a role in the realization of the action. As Jorba and Vicenteargue, if the function of inner speech is to put on a propositional content in a format that allows our 'inner eye' to access the content, then the format theory explains how we can produce a metarepresentational state, e.g. 'I believe that P,' from an utterance with the content P.[15] However, if the outcome of the cognitive processes involving inner speech episodes are second-order states, it is difficult to see how they can affect the first order states that, after all, are the producers of the behavior at stake. As Martínez-Manrique and Vicente say:

> [T]he model they propose seems to only be able to explain how IS gives us knowledge of what and how we think. Let's say that by using sentences of our language, we are able to have some kind of object before our minds. What do we

[15] Jorba and Vicente, "Cognitive phenomenology, access to contents, and inner speech;" see also, Martínez-Manrique and Vicente, "The Activity View of Inner Speech."

> gain with that? Presumably, we only gain knowledge about what we are thinking. We "see" the sentence, get its meaning, and reach the conclusion "ok, I'm thinking that p." This knowledge about what and how we are thinking may be very useful, of course, but we would say that this is only a use of IS, among many others. The account, in any case, does not explain how thought-contents are made access-conscious.[16]

That is to say, gaining access to our mental states by producing a self-ascribed metarepresentational state does not account for how our actions or the first-order mechanism are monitored, evaluated or regulated. Furthermore, the format view does not seem to respect the way we experience the inner speech episodes. When our physicist talks to herself 'the dinner!' or 'I should go to the grocery store,' she is encouraging herself to perform the action in the same way she would do it when directing these sentences to someone else. In this sense, the type of experience associated with the inner speech act is analogous to the external speech act but it does not seem to bear any resemblance with our ascriptions of mental states as the emphasis on the metarepresentational aspects suggests. In this sense, the format view does not respect our intuitions regarding how we experience inner speech episodes.

Certainly, the defenders of the format view could exploit other argumentative strategy. For instance, defending that the inner speech episodes that lead to self-ascriptions of the type 'I believe that P' or 'I desire that P' play a decisive role for a special kind of metacognition: future directed self-control. Future directed self-control requires evaluating our mental states and explore the type of genuine actions and processes that derivate from these ascriptions. In this sense, the defender of the format view could attribute to inner speech some kind of cognitive control over the behavioral consequences of their past, present and future mental states. Vierkant has illustrated this move through an example of Parfit where a young communist wins the lottery.[17] The young communist knows that rich people uses to be conservative, so he considers that if he does not get rid of the money (donating), he will become someone who does not want to be in the future, a conservative. So, the young communist is in the difficult position of donating the money and stick her ideals, or enjoying a comfortable life but becoming someone that he now would detest. The kind of mental skills the young communist engages in his considerations require self-ascribing mental states to

[16] Martínez-Manrique and Vicente, "The Activity View of Inner Speech," 4-5.

[17] See Tillman Vierkant, "What metarepresentation is for," in *The foundations of metacognition,* eds. Michael Beran, Johannes Brandl, Josef Perner, and Joëlle Proust (Oxford: Oxford University Press, 2012). The example appears in Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

himself and his future self, that is, metarepresentations. In this view, the defenders of the format view can embrace the idea that inner speech, as a producer of metarepresentations, will allow the young communist to attribute mental states to himself and his future self in order to evaluate which pattern of action to follow in the present given his attributions. This and analogous cases, where metacognitive capacities involve self-ascriptions, seem to be a plausible way for resisting the onslaughts against the format view.

This brings me to the second objection. Notice that the rationale for the format view is that inner speech facilitates the detection of underlying mental states that, after being metarepresented, we can manipulate. This idea assumes that our inner speech episodes voice or express the causally efficacious mental states that compose our first-order processes. However, this idea conflicts with empirical evidence regarding the phenomena of confabulation.[18] These studies show that humans are not always aware of the real causes of their actions, and in fact, they systematically provide ad hoc reasons to rationalize them. For instance, in the classic experiments carried out by Nisbett and Wilson, several subjects were asked which pair of panties they prefer and why. The panties were distributed on a table in a way that the subjects chose them by the distribution but they appeal to aspects such as the elasticity and the quality even though the panties were the same. These and other studies speak in favor of the idea that our reasons often are an instance of confabulation. Following this reasoning, it expectable to assume that our inner speech episodes do not necessarily voice our real mental states, and thus, it would be problematic to assume that the mental states the young communist attribute to his present self really reflect his mental states. Likewise, it is not clear how the mental states he ascribes to himself were real descriptions of his current mental states, and thus, played a causal role to modify his behavior for non-ending up being a conservative old person.[19]

Furthermore, even accepting the format view as an accurate explanation of this kind of metacognitive control, the explanatory power of the theory is too

[18] Richard Nisbett and Timothy D. Wilson, "Telling more than we can know: Verbal reports on mental processes," *Psychological review* 84, 3 (1977): 231; Michael Gazzaniga, *The mind's past* (Berkeley: University of California Press, 1998); Timothy D. Wilson, *Strangers to ourselves* (Cambridge: Belknap. 2002); Thalia Wheatley, "Everyday confabulation," In *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*, ed. William Hirstein (New York: Oxford University Press, 2009).

[19] Admittedly, not all versions of metacognition as access necessarily have troubles for explaining confabulation. An instance of this is Peter Carruthers, *The opacity of mind: an integrative theory of self-knowledge* (New York: Oxford University Press, 2011). However, they still would have to answer the telepath argument. Thanks to Tobias Störzinger for bringing my attention to this.

restricted. Although the cognitive function the young communist exercises could be accurately captured by the format view, the explanatory power of the theory is restricted to the cases involving metarepresentations, leaving aside cases where we directly control our first-order processes and behavior without such metarepresentations. As a conclusion, the format view cannot give a satisfactory explanation of how inner speech, as facilitator of second-order access, provides the acquisition of metacognitive capacities that regulate, modify or evaluate our cognition and behavior.

4. A Commitment-Based Approach to Inner Speech

In the previous section, I offered several arguments against the format view of inner speech. The aim of this section is to provide an alternative to the format view. This alternative, known as commitment-based approach, has been recently proposed by Geurts as an appropriate understanding of the cognitive functions of inner speech.[20] For the purpose of this article, I concentrate on how this approach can convincingly account for the role of inner speech in metacognition.

The commitment-based approach starts from the idea that the functions of inner speech derivate from the functions that speech acts play in coordinating agents in social interactions.[21] One way to capture how speech acts facilitate coordination between agents is by attending to how they modify the normative statuses of the speakers and her audience in terms of the commitments, duties and enabling conditions the speaker and audience undertake.[22] For instance, Geurts presents the idea as follows: "Commitment is a sine qua non for action coordination: social agents must rely on each other to act in some ways and refrain from acting in others. Commitments are coordination devices, and the main purpose of communication is to establish commitments."[23] Similarly, Kukla and Lance understand speech acts in terms of pragmatic input and outputs, where the

[20] Geurts, "Making sense of self talk."

[21] The idea that the function of inner speech derivates from the social function of outer speech is often traced back to Lev S. Vygotsky, *Thought and language*. For contemporary versions of these idea see Martínez-Manrique and Vicente, "The activity view of inner speech;" Jorba and Vicente, "Cognitive phenomenology, access to contents, and inner speech,"Fernández Castro, "Inner Speech in Action;" or Geurts, "Making sense of self talk."

[22] Robert Brandom, *Making it explicit: Reasoning, representing, and discursive commitment* (Cambridge: Harvard university press, 1998); Rebecca Kukla and Mark Norris Lance, *'Yo!'and'Lo!': The Pragmatic Topography of the Space of Reasons* (Cambridge: Harvard University Press, 2009), Bart Geurts, "Communication as commitment sharing: speech acts, implicatures, common ground," *Theoretical linguistics*(2019).

[23] Geurts, "Making sense of self talk," 8.

inputs are a set of enabling conditions and the outputs are a set of commitments, duties and entitlement the speaker and the audience undertake when recognizing the force and content of the speech act. In these views, the commitments we undertook when performing a speech act can be seen in terms of new possibilities for action. For instance, If I promise to someone that I will go with her to the theater, I am expressing a set of commitments with particular patterns of actions, including being at the theater at the time we stipulate. Certainly, not all speech acts present these direct goal-oriented commitments but even when one performs an assertion, the speaker is exhibiting certain commitment with what is rationally and socially expected from this assertion. For example, if I assert that the ice of the lake is dangerously thin, I am committing myself with future patterns of actions my audience is entitled to expect: that I will not skate on the ice or that I will warn other people of the danger. In other words, asserting something is expressing certain commitments with actions that our audience may expect us to follow.

Notice that carrying out a speech act does not necessarily involve we are in a particular mental state. As Geurts puts it:

> Commitments are obligations, and although they may be underwritten by suitable mental states, it is not necessary that they are. Insincere commitments are as binding as sincere ones, and there are unintended commitments, too. If I raise my hand at an auction, I thereby commit myself to be making a bid for whatever is currently under the hammer, even if I have no intention of doing so. True, I can try to get out of my commitment, for example, by arguing that I was only waving away a fly, but that presupposes there is a commitment to be undone.[24]

The patterns of actions associated with the commitments that follow from a particular speech act do not necessarily rely on the assumption that we are in particular mental state causally connected to these actions. Instead, the theory assumes that certain normative structures (rational and social) police our interactions in a way that connect the content of our commitments with such patterns of actions. For instance, we know what to expect from someone asserting that the ice is dangerously thin because we know what an agent ought to do in such circumstances given the rational and social structures that regulate our actions.

The commitment-based approach can help us to explain the social functions of our speech acts. The main advantage of this view is that it can account of the role of our speech acts in social coordination without reducing them to a mere exchange of information. Given that, the view is better posed to explain the speech acts whose function cannot be explained in terms of the information they provide

[24] Geurts, "Making sense of self talk," 9.

to the audience, that is, speech acts such as commands or promises, whose function does not seem to rely on how the audience gain certain piece of knowledge. Furthermore, the approach gives an automatic explanation of how our speech acts are connected to our social actions, so how they facilitate the coordination between speaker and audience.

This theoretical apparatus allows us to account the cognitive functions of inner speech in terms of the social functions of outer speech. That is, the inner speech episodes play a functional role in our cognitive machinery that is analogous to the role that external speech acts play in our social interactions. When someone asserts internally that ice of the lake is too thin, one is giving rise to private commitments with what is followed from the ice of the lake being too thin. So, she can regulate her actions and align her mental states in accordance with the commitments associated with the content of the assertion. Similarly, when an agent privately commands something to herself go to the store, she gives rise to certain goal-directed commitment to perform the action of going toward the store.

At this point, one may object that there is an important disanaology between outer and inner speech. Notice that, according to the commitment-based approach, the main function of communication is to coordinate agents. However, it is not entirely clear what exactly is the analog to coordination in the case of self-talk. In other words, if the function of inner speech derivate from the coordinating role of outer speech, then there must be a clear analog for coordination in the inner case. In order to address this challenge, one may argue that the function of outer speech for coordinating agents lies on the entitlements and commitments our speech instantiates. Once we learn how outer speech are associated to different patterns of action and cognition via those commitments and entitlement, we can rehearse such episodes in order to trigger the appropriate patterns.[25]

[25] Further, one may argue that, as for the case of intentions, inner speech episodes, as prompters of commitments, can promote intra-personal coordination by aligning volitional attitudes and practical reasoning. For instance, if I say to myself 'I will take the bus earlier tomorrow', this episode can instantiate a commitment that will help me to align my desire-like attitude toward intending to take the bus with the practical reasoning capacities necessary to find the more rational way to perform the action. For such a view regarding intentions see Michael Bratman, *Intention, plans, and practical reason* (Cambridge, MA: Harvard University Press, 1987) and Elisabeth Pacherie, "Conscious Intentions: The Social Creation Myth," *Open MIND* 29 (2015).

5. The Metacognitive Functions of Inner Speech

This section aims to account for the metacognitive functions associated with inner speech without postulating second-order access mechanisms o metarepresentational capacities. To see the contrast, notice that the format view appeals to the representational information included in the inner speech episode that produces a metarepresentation of the agent being in certain mental state in order to explain cognitive and behavioral flexibility. As I argued before, the two fundamental problems of this view are that self-ascriptions do not necessarily involve the capacity of modifying our first-order mental processes and actions. So, we can conceive circumstances where an agent ascribes to himself a particular mental state but this ascription does not make any difference. Furthermore, we can conceive several circumstances where agents regulate their actions and mental processes without having access to these states. In other words, intervening our own cognition and action do not require metarepresenting or accessing our mental states.

In order to see how the commitment-based approach can explain the connection between inner speech and metacognition, consider again the example of the physicist explained in section 3. The physicist privately utters the expression 'the dinner' which make her remember she has a dinner that night. Furthermore, she says to herself 'I should go to the grocery store' after considering she did not have food at home. The rationale behind the idea that the action of the physicist exhibits a kind of metacognitive endeavor rely on the fact that she refrains to perform the action she was doing (going back home) and triggers a new action on the light of new considerations. In this sense, she evaluates the situation and regulates her cognitive mechanisms to change her mind and carry out the action of going to the store instead of going home. The problem of the format view is that the outcome of the physicist's chain of reasoning is a self-ascription that in principle does not necessarily involve to regulate her action. Furthermore, it is hard to see how we can understand her regulatory capacities in terms of access to a mental state, especially when her private episode 'I should go to the store' does not seem to be a previous mental state in the physicist cognitive machinery, rather than a conclusion she has arrived from an episode of reasoning considering the situation. Given that, the format view should accept that the mental state represented by the private speech 'I should go to the store' was previously instantiated in the physicist's mind or abandon the idea that this case represents a case of metacognition in terms of access.

In the commitment-based approach, we can account for the case of the physicist in terms of evaluation and regulation. The metacognitive capacities

displayed has to do with evaluating an action or mental processes in accordance with certain commitments and regulating first-order mental processes and patterns of action to align them with these commitments. When the physicist brings into consciousness her memory episode through the expression 'the dinner' she evaluates her current actions in terms of the commitments the utterance expresses. Thus, she refrains to go back home when considering her utterance gives rise to certain commitments her current action is not instantiating. In other words, her current action was not conforming the expected patterns given the restrictions

imposed by the commitments of having a dinner that night. On the other hand, when she concludes that she should go to the store, she is privately committing herself with the appropriate pattern of action, and thus, she can regulate her actions in accordance with such commitments. In this sense, the inner utterance expresses the same set of commitment with actions that the sentence will express when used in a conversation with the purpose of coordinating with another person.

This position differs from the format view in two fundamental aspects. Firstly, metacognition is associated with the notions of evaluation and conformation, rather than to the notion of access. When we assert P privately, we express a set of commitments that draw a cognitive trajectory we tend to conform in order to perform what these commitments prescribe us to do, that is, self-imposed constraints to our actions. In this sense, the commitment-based approach allows us to account for the metacognitive function of inner speech in terms of evaluation and conformation rather than in terms of access. Following Proust's idea, the type of cognitive and behavioral flexibility associated with metacognition does not require the agent to access her own mental states. In my view, rather revealing our previous mental states, our metacognitive capacities shape our cognition and action by triggering different prospective patterns we are inclined to follow given the commitments that the private episodes of inner speech generate.[26]

Secondly, respecting our intuitions, the metacognitive function of inner speech is not related to the notion of metarepresentation. Modifying our cognitive capacities in a flexible way does not require being able to self-ascribe mental states. In several occasions, the regulation or evaluation of our cognition and action do not require engaging in metarepresentational thinking. In fact, we often engage in reasoning chains that lead us to a private judgment that we do not hold before, and thus, do not represent previous mental states. When we arrive at these judgments we can modify or regulate our actions in the light of the commitments these judgments without the necessity of self-ascribe any particular mental state. In

[26] Proust, *The philosophy of metacognition,* 53-78.

other words, the effective power of the inner sentence to instantiate the appropriate pattern of action does not require the person to be in the state associated with the sentence, and far less, to represent such mental states.

6. Objections

In the previous sections, I have offered a theoretical model of inner speech that account for some metacognitive functions without appealing to metarepresentations or taking for granted that metacognitive capacities require accessing mental states. This move enables us to get around the concern of the format view of inner speech. However, one may wonder whether or not embracing the commitment-based approach could give rise to another type of problems. In principle, there are two main objections one may envisage for the commitment-based approach. Firstly, one may argue that future directed self-control (see section 3) fall out of the explanatory reach of the commitment-based approach. Secondly, one may consider that the notion of speech acts in terms of commitments is problematic or, at least, unnecessary for explaining the function of inner speech. This section is devoted to addressing these two objections.

For addressing the first problem, consider again the case of the young rich communist. As we have seen, Vierkant argues this case exemplify a kind of metacognitive capacity that cannot be performed without the metarepresentations and access required by the format view. Given that, one may wonder whether this kind of metacognitive control is a feasible counterexample against the commitment- based approach. After all, the young rich communist case exhibits the features of metacognitive control the commitment-based approach casts into question as necessary for the display of the metacognitive function of inner speech. Now, it must be clarified that the commitment-based approach is compatible with the fact that we can display mental concepts (belief, desire, fear) in our reasoning or inner speech episodes. In fact, we often self-attribute mental states (avowals) putting those mental concepts into work. However, this does not mean such self-ascriptions endow us with a particular mental access to our own psychological states.

In fact, when we pay closer attention to the social role of self-ascriptions, we realize that in conversational contexts we often use the first-personal ascription with pragmatic purposes.[27] For instance, the phrase 'I think' is frequently presented

[27] James O. Urmson, "Parenthetical verbs,"*Mind* 6, 244 (1952): 480–496; Karin Aijmer, "I think: an English modal particle," in *Modality in Germanic Language: Historical and Comparative Perspectives*, eds. Toril Swan and Olaf Westik (De Gruyter Mouton, 1997); Anna Wierzbicka, *English: Meaning and culture* (Oxford: Oxford University Press, 2006); Mandy

as having the function to mitigate the degree of commitment to the sentence it ranges. Wierzbicka provides a deep analysis of parenthetical uses of 'believe,' 'think' and other mental verbs. She claims that the verb 'think' conveys the meaning of disclaiming knowledge "not by saying "I don't know" but by saying "I don't say: I know it."[28] In other words, 'I think P' expresses a certain degree of caution. Similarly, the verb 'believe' (in contrast to 'I think' for instance) seems to play an indicative function. As Aijmer claims: "I believe does not only express a subjective attitude. It also conveys that the speaker has some evidence for what he says."[29] We can see the contrast between 'I think' and 'I believe' in the incompatibility of 'I believe' with phrases like 'I'm not sure.' While 'I think that Riga is the capital of Latvia, but I'm not sure' is idiomatic, 'I believe that Riga is the capital of Latvia but I'm not sure' is not. This difference between the level of reliability that 'think' and 'believe' convey must not divert our attention away from the fact they share their basic function: they are devices for canceling or altering the speaker's commitments. The verbs 'believe' and 'think' seem to be mitigators of the force of the claim. Of course, parenthetical uses are not restricted to these types of indications involving mitigations. Verbs as 'rejoice' or 'regret' indicate emotional orientation, others as 'wish' or 'desire' indicate the preference toward the commitments of the statement. What these parenthetical uses of propositional attitude verbs share is its function for providing indications or prescriptions to the hearer about how to evaluate the commitments of the proposition associated with the mental verb. As a conclusion, mental verbs in self-ascriptions seem to have the pragmatic function of signaling certain attitudes or indications toward the commitments expressed by the statement under the scope of the mental verb.

Taking this inside on board, when the young rich communist is evaluating what to do in the light of his future belief 'I will believe social justice does not matter,' he is considering the commitments he will give rise in the future given the content of his future belief. Furthermore, he assesses the type of actions he must carry out in the present in order to avoid his future commitments with the assertion that social justice does not matter. In this sense, we can recruit the same kind of commitment-based explanation without bringing out any type of access-like explanation. Although this kind of explanation seems to necessitate certain notion of metarepresentation that allows the young rich communist to perceive

---

Simons, "Observations on embedding verbs, evidentiality, and presupposition," *Lingua 117*, 6 (2007), 1034-1056.

[28] Wierzbicka, *English*, 38

[29] Aijmer, "I think," 17

himself as a minded creature, it does not commit us with understanding self-ascriptions as descriptions of inner processes or psychological states, rather than expressions that make explicit the commitments with the present and future actions associated with the content of the proposition under the scope of the mental verb. Thus, the commitment-based approach could also give a plausible explanation of the metacognitive capacities the future directed self-control requires.

A second objection against the commitment-based approach may cast into question its plausibility as a theory of the social function of speech acts. One may argue, for instance, that a neo-Gricean model of communicationprovides a better understanding of communication, and subsequently, for the cognitive function of inner speech.[30] In the neo-Gricean model, a hearer expects certain patterns of actions from a speaker because her speech acts express certain mental states that are causally connected with the given action. For instance, when a speaker asserts P, the hearer can infer through different pragmatic mechanisms that he is expressing a belief that P, and thus, the hearer can expect from the speaker a range of patterns of actions causally connected with such belief. The neo-Gricean approaches to communication exhibit certain problems whose consideration is beyond the purpose of this paper.[31] However, for the purpose of this article, it is sufficient to notice that such position requires our speech act to voice certain underlying mental states, which again brings out the problem of confabulation. Considering that inner speech requires putting to work pragmatic mechanisms that infer the mental states of the agent implies that the agent must be in a particular mental state that is causally connected to the private episode. However, as the empirical evidence considered in section 3 emphasizes, it is problematic to assume that our reasons, and thus our inner speech episodes always reflect an underlying mental state.

On the contrary, this is not problematic for the commitment-based approach. As Strijbos& de Bruin argue, our confabulatory reasons can have two

[30] For two well-known neo-Gricean Models of communication see Kent Bach and Robert Harnish, *Linguistic Communication and Speech Acts* (Cambridge, Mass.: MIT Press, 1979); and Dan Sperber and Deidre Wilson, *Relevance: Communication and Cognition*, (Oxford: Blackwell, 1986).

[31] For instance, these approaches are usually committed with the idea that communication requires the instantiation of mindreading mechanisms that, as Tadeusz Zawidzki has emphasized, make mental state attribution computationally intractable (see Tadeusz Zawidzki "The function of folk psychology: Mindreading or mindshaping?" *Philosophical Explorations* 11, 3 (2008): 193-210).

purposes.[32] Firstly, they can help to give coherence to our previous actions by providing us with a narrative. Secondly, they can have a prospective function, generating commitments that we are inclined to conform, and thus, that regulate our behavior and cognitive mechanisms. In this sense, the commitment-based approach can help us to elucidate the regulatory function of inner speech while avoiding the problem of confabulation. That is, our inner speech episodes do not necessarily reflect our underlying mental states, rather than it help us to give coherence and regulate our actions by giving rise to the commitments with certain patterns of actions.

7. Conclusion

The aim of this paper was to present several concerns regarding the format view of the metacognitive capacities of inner speech and to advocate an alternative. The problems associated with the format view rely on the role that the model assigns to metarepresentations and the notion of access. The solution I have offered respects our intuitions concerning inner speech episodes and accounts for the metacognitive capacities of regulating and evaluation our cognition and action. This position offers an alternative that does not require postulating metarepresentations or considering thinking about thinking in terms of access. Furthermore, the theory can avoid two possible objections. On the one hand, it can account for the cases where our metacognitive capacities require self-ascriptions. On the other hand, the theory can avoid certain challenges that other views of communication that have enjoyed a greater popularity cannot avoid.[33,34]

[32] Derek Strijbos and Leon de Bruin, "Self-interpretation as first-person mindshaping: implications for confabulation research," *Ethical Theory and Moral Practice* 18, 2 (2005): 297-30.

[33] This article was written thanks to the funding provided by the project "Inner speech, Metacognition, and the Narrative View of Identity" (FFI2015-65953-P)and "Contemporary Expressivism and the Indispensability of Normative Vocabulary" (FFI2016-80088-P) funded by Ministerio de Economía y Competitividad; and the postdoctoral research contract "Puente," funded by the University of Granada.

[34] Thanks to all members of the Department of Philosophy at University of Granada (AKA Granada Gang) for helpful comments on previous versions of these paper. I would like to also to acknowledge the influence and support of Fernando Martínez-Manrique and Agustin Vicente.