teorema

Vol. XXXIV/1, 2015, pp. 135-148 ISSN: 0210-1602 [BIBLID 0210-1602 (2015) 34:1; pp. 135-148]

Are Alien Thoughts Beliefs?

Lisa Bortolotti and Kengo Miyazono

INTRODUCTION

Thought insertion is a common delusion in schizophrenia. People affected by it report that there are thoughts in their head that have been inserted by a third party. These thoughts are self-generated but subjectively experienced as alien (hereafter, we shall call them *alien thoughts* for convenience). In chapter 5 of *Transparent Minds*, Jordi Fernández convincingly argues that the phenomenon of thought insertion can be explained as a pathology of self-knowledge. In particular, he argues that the application of the bypass model of self-knowledge can shed light on what is amiss in people who experience thought insertion.

In this brief commentary, we examine the proposal by Fernández, highlight some of its strengths, and raise one main objection to it. In mainstream philosophical accounts of thought insertion, people who report the delusion are thought to have ownership of the alien thoughts, but to lack a sense of agency with respect to such thoughts, and this is supposed to explain why they ascribe the thoughts to someone else. Fernández correctly identifies the limitations of mainstream accounts of thought insertion. First, to claim that people with the delusion of thought insertion own the alien thoughts does not sit well with the phenomenology of thought insertion as it is expressed in people's self-reports. Second, people do not need to experience a sense of agency with respect to a thought in order to ascribe the thought to themselves. Sense of agency is too demanding a condition for self-ascription.

Fernández puts forward a novel account of thought insertion, attempting to solve the problems identified with mainstream accounts. He defends the view that thought insertion is a failure of self-knowledge that consists in the person failing to ascribe one of her be-

liefs to herself. Similar accounts to the one Fernández develops in the chapter have been defended by Bortolotti and Broome (2009) and Pickard (2010). For Fernández, the failure in self-ascription is due to the fact that the person experiencing thought insertion does not feel pressured to endorse the content of a belief that she has [Fernández (2013), p. 142]. Fernández explains this phenomenon in detail by reference to the bypass model of self-knowledge.

When Fernández develops his proposal, he relies on the claim that the alien thought is a belief. We challenge this claim. We agree with Fernández that, in thought insertion, the person does not ascribe the alien thought to herself and does not commit to the content of the thought being true, even though the thought is self-generated. Such considerations suggest to us that the alien thought is not a belief. Given the elusive nature of mental states in general and beliefs in particular, this may seem just a terminological issue. But in the present context, whether the alien thought is a belief is important in order to offer a characterisation of thought insertion that does justice to the first-person reports and other behavioural manifestations of people with the delusion. One may also argue that, as the model of selfknowledge Fernández applies to alien thoughts (the bypass model) is a model of self-knowledge for beliefs, such a model could only feature in the explanation of thought insertion if alien thoughts were beliefs. But we do not think this is necessarily the case, because, in the book, Fernández successfully applies his bypass model to other types of mental states, such as desires.

I. COMPETING ACCOUNTS OF THOUGHT INSERTION

Typically, people with thought insertion report that a third party has inserted a thought into their mind. This is a curious thing to report, because usually people can only directly access self-ascribed thoughts. In thought insertion, the content of a thought is accessed directly ("from the inside"), but is not self-ascribed. What makes the phenomenon so hard to describe is that there is no theory-neutral characterisation of it, apart from the claim that people have thoughts they do not feel are their own and that they attribute to someone else. Different philosophers interpret and explain thought insertion differently, and offer different answers to the following questions: *Do people with*

thought insertion own the alien thoughts? Do they experience a sense of agency towards the alien thoughts? Do they endorse the content of the alien thoughts?

The very meaning of the three questions above is subject to interpretation in the philosophical literature, and one obvious example is the notion of ownership of thoughts. Some endorse a weak conception of ownership of thought, according to which it is sufficient to be directly aware of the content of a thought and to be able to locate it in one's own mind in order to own it [Gallagher (2004)]. Other authors opt for a more demanding notion of ownership, according to which the thought needs to be self-ascribed in order to be owned [Bortolotti and Broome (2009)]. Fernández rejects weak notions of ownership of thoughts and assumes that ownership implies self-ascription. In this demanding sense of ownership, people with thought insertion do not own the alien thoughts.

Philosophers also disagree about whether one can be the agent of a thought (or have a sense of agency with respect to a thought) in the same way that one can be the agent of a physical movement (or have a sense of agency with respect to a physical movement). Usually, agency implies having initiated the thought or the movement. A person has a sense of agency with respect to a thought if she feels that she has produced, or she has been causally responsible for, that thought. People with thought insertion have no sense of agency with respect to the alien thoughts, but the concern is that the sense of agency may also be missing from other (non-pathological) experiences of thoughts or beliefs. After all, unsolicited thoughts or even beliefs can "pop up" in people's minds without generating delusional hypotheses. Fernández rejects the view that people with thought insertion fail to self-ascribe the alien thoughts because they have no sense of agency with respect to those thoughts. Failed sense of agency seems compatible with selfascription in some cases, and thus we need an alternative explanation for the peculiarity of the experience of thought insertion.

The rejection of the no-sense-of-agency account motivates Fernández's proposal. According to the no-sense-of-agency account, the person with thought insertion is directly aware of the content of the alien thought, but she fails to experience a sense of agency with respect to it [Gerrans (2001); Gallagher (2004); Graham and Stevens (2000); Campbell (1999) and (2002)]. There are at least two main versions of the no-sense-of-agency account (one based on a *personal* explanation

of agency and the other on a sub-personal one), and both share the same problem. They start from the assumption that people standardly experience a sense of agency with respect to their thoughts, and that people with thought insertion do not. More specifically, the core of the account is that the experience of thought insertion is the experience of not having a sense of agency with respect to one's thoughts. But this does not seem right. No matter how the sense of agency is characterised (as the act of thinking requiring deliberation and effort, or as the thought being driven by explicit intentions), this is not an experience people necessarily have when they have thoughts that they are happy to ascribe to themselves. An alternative to the no-sense-of-agency account of thought insertion is the no-authorship account [Bortolotti and Broome (2009)]. According to the no-authorship account, the key issue is not whether the person feels that she is the agent of the thought, but whether she is genuinely committed to the content of the thought being true. We take important aspects of this approach to be shared by Fernández (2013) and Pickard (2010), although terminology and more substantial details differ considerably from one version of the approach to the next.

II. THE BYPASS MODEL

Fernández's bypass model is about how people determine that they have a belief from a first-personal point of view. There are two main features of the bypass model. The first feature is *non-neutrality*. If Helen determines that one of her beliefs is that Madrid is in Spain, then that belief is "presented as being correct" to her, and she does not maintain a neutral attitude towards the claim that Madrid is in Spain [Fernández (2013), p. 167]. Another way to capture the phenomenon is to say that Helen identifies with her belief.

If I determine, from a first-person point of view, that one of my beliefs is that my wife is cheating on me, then that belief is not presented to me as being neutral on whether she is actually cheating on me or not. That belief is presented to me as being correct. [...] I am compelled to endorse the content of that first-person belief [Fernández (2013), p. 167].

The second feature is *transparency*. To ascribe the belief that Madrid is in Spain to herself, Helen does not need to *look inward*, and scrutinise

her own beliefs, but she needs to *look outward*, and ask how things are (e.g., whether Madrid is in Spain). The main idea is that the ground for Helen's second-order belief that she has the belief that Madrid is in Spain is the same ground for the belief that Madrid is in Spain.

When I form the belief that I have a certain belief, the state on the basis of which I form my higher-order belief normally constitutes grounds for the first-order belief in me. What I do is to look, as it were, past the belief in order to self-attribute it [Fernández (2013), p. 49].

The notion of identification (as non-neutrality) and the notion of transparency are at the core of Fernández's theory of how people determine what they believe, and we feel that these are features of the way people know about their beliefs, not necessarily features of the way people know about their other mental states. For instance, people "look past" beliefs but do not necessarily look past other mental states. Let us briefly consider imaginings as a contrast case. Helen's belief that she imagines Madrid being in England is not grounded in Madrid being in England. But Helen's belief that she believes that Madrid is in Spain is grounded in Madrid being in Spain. Beliefs are transparent, but imaginings are opaque. When Helen determines that she imagined that Madrid was in England, she does not need to feel any compulsion to endorse that Madrid is in England. As we saw, when Helen determines that she believes that Madrid is in Spain, the belief that Madrid is in Spain is presented to her as correct. Beliefs about beliefs imply commitment, but beliefs about imaginings do not. According to Fernández, the person with thought insertion does not have the same experience as Helen when she determines that she believes that Madrid is in Spain. The person with thought insertion experiences neutrality towards the content of one of her beliefs, and is not compelled to revise the belief or abandon it when evidence against it emerges. It is as if the belief were other than a belief (e.g., an act of imagination).

When the patient is aware of those beliefs that she disowns, she does not experience that those beliefs are forcing her to accept any particular picture of the world. That is, she does not feel compelled to endorse their contents [Fernández (2013), p. 197].

In this section, we saw that two key aspects of Fernández's explanation of thought insertion rely on the alien thoughts being beliefs.

Why? Because Fernández identifies the problem of thought insertion with people failing to experience something they ought to experience when they determine from a first-person point of view that they have a certain belief, i.e., non-neutrality and transparency. People with thought insertion experience some of their beliefs as neutral in content, and opaque. If alien thoughts were not beliefs, then neutrality and opacity might not be a problem.

In the rest of the paper, we shall assess Fernández's own account in light of two of his methodological requirements.

III. MATCHING FIRST-PERSON REPORTS

As we saw, Fernández argues that people with thought insertion are aware of their alien thoughts (e.g., "The grass is green"), but cannot ascribe such thoughts to themselves via the bypass model, and thus they do not recognise them as their own beliefs. The higher-order belief (e.g., "I believe that I believe that the grass is green") does not have the *assertive phenomenology* that higher-order beliefs of this sort usually have. Thus, people with thought insertion do not feel *pressured* to endorse the content of the alien thoughts. They have introspective access to their thoughts without identifying with them. This is a pathological form of self-knowledge.

The application of the bypass model to thought insertion seems problematic in a number of ways. First, the phenomenology of thought insertion as captured by first-person accounts is not compatible with alien thoughts being beliefs. Second, Fernández's main argument for regarding alien thoughts as beliefs, that their being beliefs explains the puzzlement in people's reports of thought insertion, can be easily challenged.

Fernández stipulates that any account of thought insertion should respect the first-person reports of the experience: "Any account of thought insertion must respect the patients' own reports" [Fernández (2013), p. 149].

We agree that this is a sensible requirement, and is too often neglected in the philosophical literature, for instance when philosophers claim that people *own* alien thoughts. But we are not convinced that Fernández's proposal satisfies this requirement. People do not describe alien thoughts as their beliefs or as beliefs at all, but Fernández

argues that they are beliefs and his account heavily relies on them being beliefs.

When we read the reports of thought insertion listed by Fernández, we have the impression that people do not self-ascribe the alien thoughts as beliefs or as any other mental state. People report the thoughts, but judge them to be alien and offer an explanation of their being in their head which is implausible, — e.g., the thoughts come from somewhere or are someone else's and have been inserted in the person's head for some purpose. The thought that is being regarded as alien (e.g., "The grass is green") is never described as a belief. The delusional explanation of that thought being in the person's head (e.g., "My neighbour has inserted the thought that the grass is green into my mind") is endorsed and defended as a belief.

For Fernández the explanandum is the phenomenon of awareness without ownership [Fernández (2013), pp. 144-5, also see footnote 11]. This is defined as follows.

For any subject with thought insertion S, there is some belief B such that:

- (i) S claims to have B;
- (ii) S claims that B is not her belief.

The author identifies B as a belief, and the claims made by the subject are interpreted as stating that (i) the subject has a belief and that (ii) the belief is not her belief. Conditions (i) and (ii) are problematic; in none of the first-person accounts of thought insertion listed by Fernández can subjects be interpreted as saying: "I have a belief that is not mine". Not only do people refrain from talking about *believing* or *having beliefs* but, with one exception (Patient 7), they do not even say that they *have thoughts*. They say that *thoughts come to them*, that they are *given to them* or *put into their mind* [Ibid. pp. 143-144].

Patient 1: The thoughts of Eamonn Andrews come into my mind.

Patient 2: Thoughts come into my head.

Patient 3: It's like a thought as it comes in.

Patient 4: They were being put into my mind.

Patient 5: They come unasked. They come at any moment like a gift.

Patient 6: The thought was given to me.

Patient 7: The words just *came into my head*. They were ideas I was having.

The first-person reports that Fernández appeals to do not seem to support his hypothesis that alien thoughts are beliefs. In some of them, alien thoughts are more like commands rather than beliefs. In other reports, alien thoughts seem to be imagistic or auditory rather than doxastic. Fernández says that it is intuitive that alien thoughts in Patient 1 and Patient 7 are beliefs. But Patient 1 says: "He treats my mind like a screen and flashes thoughts onto it like you flash a picture". And Patient 7 uses terms such as "hear" and "words" in describing the inserted thoughts. In addition, there are some reports that Fernández does not cite where the alien thoughts are more like desires or emotions rather than beliefs [Mellor (1970), p. 17]:

A 23-year-old female patient reported, 'I cry, tears roll down my cheeks and I look unhappy, but inside I have a cold anger because they are using me in this way, and it is not me who is unhappy, but they are projecting unhappiness onto my brain. They project upon me laughter, for no reason, and you have no idea how terrible it is to laugh and look happy and know it is not you, but their emotions.'

A 26-year-old engineer emptied the contents of a urine bottle over the ward dinner trolley. He said, 'The sudden impulse came over me that I must do it. It was not my feeling, it came into me from the X-ray department, that was why I was sent there for implants yesterday. It was nothing to do with me, they wanted it done. So I pick up the bottle and poured it in. It seemed all I could do.'

The main reason Fernández offers for the claim that alien thoughts are beliefs is that people find their experiences perplexing and there would be nothing perplexing about entertaining thoughts whose truth one is not committed to [Fernández (2013), p. 146]. Why would the experience of thought insertion be odd if the patient were just entertaining a thought without endorsing its content? What would justify the person's state of puzzlement is the fact that a *belief* is there with-

out being endorsed (what Fernández describes as the experience of neutrality towards the content of a belief).

This is the challenge: can we explain the puzzlement in the reports without assuming that the alien thoughts are beliefs? People can relate to their mental states by owning them (ascribing the mental states to themselves as mental states of a certain type) and by identifying with the mental states and endorsing their content. People own unsolicited thoughts, but do not necessarily endorse their content. People own opinions, and typically endorse their content. How do people relate to alien thoughts? They do not own them, and they do not endorse their content either. This suggests that alien thoughts are not beliefs. They are *alien* (not self-ascribed, disowned) *thoughts* (not necessarily beliefs). The puzzlement, then, does not need to come from the fact that there is a belief that is entertained in a neutral way; it can come from the fact that there is a thought that is not owned. This hypothesis does not need to commit us to the alien thought being doxastic.

What is puzzling for people is that a thought that is not owned is there, in their head, in the first place. Awareness without ownership is puzzling even if the object of awareness is not the content of a belief. It is this experience that people use as evidence for their belief that the thought is not theirs and has been inserted by another.

- Patient 1: [Eamonn Andrews] treats my mind like a screen and flashes thoughts onto it like you flash a picture.
- Patient 2: They come from this chap, Chris. They are his thoughts.
- Patient 7: The houses I was passing were sending messages to me. [The words] belonged to the houses, and the houses had put them in my head.

IV. MATCHING OTHER SURFACE FEATURES OF THOUGHT INSERTION

Fernández stipulates that: "an account of thought insertion should also respect certain aspects of the patients' behaviour, and the capacity to explain them will count as a virtue of any such account' [Fernández (2013), p. 150]. The two features of delusions he lists are: "incorrigibility" and "little weight in informing action". Incorrigibility and little weight in informing action seem too strong as labels for the phenome-

na that typically characterise delusions. Some delusions in schizophrenia are resistant to counter-evidence (this fixity is a defining feature of delusions in DSM 5) and some delusions fail to drive action in a consistent way, but both phenomena should not be exaggerated [see Bortolotti (2010)].

Leaving these wider considerations aside, does the observation of inconsistent influence on action support the bypass model endorsed by Fernández? The person's neutrality towards the content of the alien thought would be evidence for the view that thought insertion derives from a failure of the bypass model. But this is problematic. First, Fernández's account of inconsistent influence on action would only be applicable to the case of the delusion of thought insertion. However, inconsistent influence on action is allegedly seen in a wide range of delusions (from Capgras delusion to delusions of persecution) that cannot be as easily explained as pathologies of self-knowledge. If inconsistent influence on action is a general feature of delusions, then an explanation that only makes sense of the phenomenon in thought insertion may not be our best option. Accounts of the phenomenon that are applicable to all delusions [see Gallagher (2009), Bortolotti (2010), Bortolotti and Broome (2012) may be preferred.

Second, Fernández seems to be under the impression that "delusions of thought insertion often do not make a difference to the subject's behaviour" [Fernández (2013), p. 170]. But the empirical evidence at our disposal does not support the view that delusions of thought insertion lack influence on action. Clinical predictions about what people with thought insertion will do are based on the delusion being endorsed and being causally efficacious with respect to future action. For instance, people with thought insertion have been reported to wear a cap or stay indoors to prevent others from inserting thoughts into their minds [Bortolotti and Broome (2012)]. Spence reports the story of a man who screamed incessantly for days to avoid thoughts being inserted into his mind: "Why was he screaming? He did it to stop the thought interference. It was the only way. Thoughts came into his head, all the time, from outside. The screaming seemed to clear his mind, helped him regain control, but then he'd need to scream again" [Spence (1999), p. 489]. There is also some evidence that "threat/controloverride" found in delusions of persecution and thought insertion is reliable in predicting dangerousness. People with no previous history of violence may become aggressive towards other people they perceive

as intruding or threatening [see Mullins and Spence (2003) for a review]. This literature suggests that the delusion of thought insertion is acted upon.

What Fernández should say (and he does say so when he discusses the relative merits and problems of the idealist conception of thought insertion) is that, whilst delusions of thought insertion can make a difference to the subject's behaviour, the *alien thought* is not acted upon. Fernández's own explanation for why the alien thought does not influence behaviour is that it is a belief the person does not identify with, a belief whose content the person feels neutral about. But another reason why people may not act on their alien thoughts is that they do not take them to be beliefs at all. A thought does not need to have a tight connection with action. Moreover, an alien thought is not presented as a belief the person has, but as an intrusion. People do not necessarily identify with it and do not feel compelled to act on it. Fernández appeals to the fact that alien thoughts do not influence action to support the application of the bypass model to thought insertion: they are beliefs whose content feels neutral. But a simpler explanation for why the alien thought is not acted upon is that it is not a belief.

CONCLUSION

In this commentary we considered Fernández's account of thought insertion defended in *Transparent Minds*. We agree with Fernández that the no-sense-of-agency account of thought insertion fails to provide an adequate explanation of the phenomenon, and we find the general core of Fernández's account persuasive and insightful: thought insertion is a pathology of self-knowledge in which the person fails to identify with the content of the alien thought.

However, we believe that Fernández is not always clear about the distinction between the alien thought and the delusion of thought insertion. We acknowledge that the delusion of thought insertion is a belief (self-ascribed, endorsed, and acted upon), but we resist the claim that the alien thought is a belief. The alien thought is not self-ascribed, not endorsed, and not consistently acted upon. The experience of awareness without ownership of the alien thoughts is sufficient by itself to account for two facts that Fernández finds interesting about

thought insertion: (1) that people find the experience of alien thoughts puzzling, and (2) that they do not act on their alien thoughts. People find alien thoughts perplexing and do not act on them because they do not ascribe those thoughts to themselves.

But how is the experience of awareness without ownership to be explained? We would prefer it if an account of the phenomenon of thought insertion did not rely on the claim that alien thoughts are beliefs, as the language of first-person reports and the other behavioural manifestations of the delusion do not support the reading of alien thoughts as beliefs. It seems to us that Fernández can provide a satisfactory account of thought insertion that respects first-person reports and behavioural features of the delusion without giving up on the bypass model. After all, the bypass model can be applied to mental states other than beliefs, and Fernández himself applies it to desires in his book. It is not necessary to commit to the doxastic nature of alien thoughts for the bypass model to contribute to an explanation of thought insertion.

Philosophy Department, University of Birmingham ERI Building Edgbaston, Birmingham B15 2TT, UK E-mail: l.bortolotti@bham.ac.uk Department of Humanities and Social Sciences Keio University 2 Chome-15-45 Mita, Minato Tokyo 108-8345, Japan E-mail: kengomiyazono@gmail.com

ACKNOWLEDGEMENTS

The authors are grateful to Matthew Broome for helpful comments.

REFERENCES

BORTOLOTTI, L. (2010), "Double Bookkeeping in Delusions: Explaining the Gap between Saying and Doing"; in K. Frankish, A. Buckareff and J. Aguilar (eds.) *New Waves in the Philosophy of Action*; Palgrave, chapter 11, pp. 237-256.

BORTOLOTTI, L. and BROOME, M. (2009), "A Role for Ownership and Authorship in the Analysis of Thought Insertion; *Phenomenology and the Cognitive Sciences*, Vol 8(2), pp. 205-224.

- BORTOLOTTI, L. and BROOME, M. (2012), "Affective Dimensions of the Phenomenon of Double Bookkeeping in Delusions"; *Emotion Review* 4 (2), pp. 187-191.
- CAMPBELL, J. (1999), "Schizophrenia, the Space of Reasons, and Thinking as a Motor Process"; *The Monist*, 82 (4), pp. 609-625.
- GALLAGHER, S. (2000), "Philosophical Conceptions of the Self: Implications for Cognitive Science"; *Trends in Cognitive Sciences*, Vol 4(1), 4, pp. 14-21.
- (2004), "Agency, Ownership, and Alien Control in Schizophrenia. Advances in Consciousness Research"; in D. Zahavi, T. Grünbaum and J. Parnas (eds.) *The Structure and Development of Self-Consciousness: Interdisciplinary Perspectives*, Amsterdam/Philadelphia, John Benjamins, pp. 89-104.
- (2009), "Delusional Realities"; in M. R. Broome and L. Bortolotti (eds.) Psychiatry as Cognitive Neuroscience: Philosophical Perspectives, Oxford, Oxford University Press, pp. 245-266.
- GERRANS, P. (2001), "Authorship and Ownership of Thoughts"; *Philosophy, Psychiatry*, & *Psychology*, 8 (2), pp. 231-237.
- MELLOR, C. S. (1970), "First Rank Symptoms of Schizophrenia"; *The British Journal of Psychiatry*, 117, pp. 15-23.
- STEPHENS, G. L. (2000), "Thought Insertion and Subjectivity"; *Philosophy, Psychiatry*, & *Psychology*, 7 (3), pp. 203-205.
- STEPHENS, G. L. and GRAHAM, G. (1994), "Self-Consciousness, Mental Agency, and the Clinical Psychopathology of Thought Insertion"; *Philosophy, Psychiatry*, & *Psychology*, 1 (1), pp. 1-10.
- (2000), When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts; Cambridge, Mass: MIT Press.

RESUMEN

En este breve comentario examinamos la explicación del fenómeno de inserción de pensamientos que da Fernández en su libro *Transparent Minds*. Subrayamos algunos de los puntos fuertes de dicha explicación, y planteamos una objeción. Según la concepción filosófica más extendida de la inserción de pensamientos, los afectados dicen tener pensamientos ajenos, pero no reconocen ser los agentes de tales pensamientos. Fernández identifica correctamente los problemas de tal concepción y propone una alternativa prometedora. Sin embargo, al hacerlo se apoya en la tesis de que los pensamientos ajenos son creencias, una tesis que es sometida a examen aquí. Los afectados por la inserción de pensamientos no se comprometen con la verdad del contenido del pensamiento ajeno.

PALABRAS CLAVE: inserción de pensamientos, delirio, esquizofrenia, autoconocimiento, posesión de pensamientos, compromiso, creencia.

ABSTRACT

In this brief commentary, we examine the account of thought insertion provided by Jordi Fernández in his book, *Transparent Minds*. We highlight some of the strengths of the account, and raise one main objection to it. In mainstream philosophical accounts of thought insertion, people who report the delusion are thought to have ownership of the alien thoughts, but to lack a sense of agency with respect to such thoughts. Fernández correctly identifies the limitations of mainstream accounts of thought insertion and articulates a promising alternative. As he does so, though, he relies on the claim that alien thoughts are beliefs, and we challenge that claim. People with thought insertion do not commit to the truth of the content of the alien thought.

KEYWORDS: Thought Insertion, Delusion, Schizophrenia, Self-Knowledge, Ownership of Thoughts, Endorsement, Belief.