

## Moore's Paradox and the Expression of Belief

*Louise Röska-Hardy*

The expression and self-attribution of belief figure prominently in the puzzles generated by what has come to be called "Moore's Paradox". G. E. Moore claims that it is absurd to assert sentences of the form

- (1) "*p*, but I do not believe that *p*" or
- (2) "I believe that *p*, but not-*p*",

despite the fact that (1) and (2) could be true. Indeed, another person could truly say of me, for example, "*p*, but she doesn't believe it" or "She believes that *p*, but not-*p*" without incurring any absurdity. The 'absurdity' pertains to utterances of (1) or (2) in the first-person singular present indicative. For I can assert (1) and (2) in other tenses, e.g. "It was raining, but I didn't believe it", without any peculiarity. Furthermore, the peculiarity does not seem to be confined to the overt assertion of (1) or (2). Any thought whose content is expressed by (1) or (2) would be equally odd.<sup>1</sup> The oddity of Moorean sentences does not seem to be due to their being *logically* self-contradictory, since both conjuncts could be true. What is odd about uttering or thinking Moorean sentences?

The question is important, because the puzzles generated by Moore's Paradox concern central issues in the philosophy of language and the philosophy of mind. Hence a satisfactory explanation of the oddity of the Paradox promises to provide insight into the character of belief and to illuminate the nature of self-knowledge.<sup>2</sup> Moreover, Robert Gordon argues that the ability to capture the peculiarity of Moore's Paradox provides a reason for favoring an ascent routine simulation approach to the self-ascription of mental states over Alvin Goldman's introspection-simulation account or a theory-theory approach.<sup>3</sup> By examining the Paradox from a semantic and from a pragmatic perspective, I want to show that several proposed explanations are inadequate. In a first step the semantic structure of sentences like (1) is examined with regard to logical form, distribution of truth values and logical consistency and it is argued that an explanation of the Paradox is not to be found at the semantic level. I then consider explanations which locate source of the Paradox at the pragmatic level, e.g. in an internal connection between the notions of belief and assertion, in speech act-pragmatic aspects of assertion, or in Gricean conditions for meaning. It is argued that none of these offer an adequate account. However, the failings of the semantic and the pragmatic explanations point to a feature of the Paradox which merits closer scrutiny.

Moore presents two versions of the Paradox, which, arguably, are not equivalent.<sup>4</sup> He introduces the first version by way of illustrating a distinction between what someone ‘says’ or ‘asserts’ and what she ‘implies’ in saying what she does. According to Moore, when I say “I went to the pictures last Tuesday” I *imply* that I believe that I went to the pictures last Tuesday, although in saying what I say I do not *assert* that I believe it, i.e. that I believe I went to the pictures “does not follow from” what I assert.<sup>5</sup> Moore traces the implied belief to the fact “that in the immense majority of cases a man who makes such an assertion ... does believe or know what he asserts. ...[T]his is why to say such a thing as ‘I went to the pictures last Tuesday, but I don’t believe that I did’ is a perfectly absurd thing to say, although what is asserted is something perfectly logically possible.”<sup>6</sup> Although it may be true both that I went to the pictures last Tuesday and that I do not now believe I did, I cannot assert it without absurdity. We are thus presented with a situation in which there seem to be truths about me which I cannot assert myself. On Moore’s view, the absurdity of asserting sentences like (1) “*p*, but I do not believe that *p*”, is explained by the belief implied in my asserting ‘*p*’. In the second version of the Paradox, (2) “I believe he has gone out, but he has not”, what is implied by my asserting ‘*p*’ is that I do not believe that *not-p*, according to Moore.<sup>7</sup>

In Moore’s view the peculiarity of asserting (1) and (2) cannot be traced to the sentence itself, since it is not self-contradictory. Neither (1) nor (2) is necessarily false. None the less, the assertive utterance of either is conceptually odd. The challenge of Moore’s Paradox is to identify the source of this oddity. Why should it be odd to assert (1) or (2) when what each sentence asserts might be true? Moore locates the source of the Paradox in the implications of assertion. On his view saying or asserting implies believing. To utter ‘*p*’ is to imply that one believes that *p*; this implied ‘I believe that *p*’ conflicts with one’s saying ‘I don’t believe that *p*’, thus generating the absurdity of such an assertion. The absurdity is due to an implied speaker’s belief which contradicts the speaker’s avowed belief. Saying that *p* implies that the speaker believes or knows that *p*. However, because ‘I believe that *p*’ is only implied and not entailed, (1) does not have to be treated as self-contradictory. A number of authors have argued that Moore’s Paradox is not limited to overt assertions.<sup>8</sup> They claim it would be just as odd to think the content expressed by (1) to be true of oneself. This suggests that an explanation of the absurdity involved in Moore’s Paradox should be applicable both to thought and to speech.

Let us consider the semantic features of Moorean sentences. Both versions of the Paradox involve sentences which are composed of the conjunction of two clauses. Consider the first version of the Paradox, “*p*, but I don’t believe that *p*”. The first clause is a (simple) present tense indicative sentence ‘*p*’ of subject-predicate form. The second clause, also of subject-predicate form, exhibits ‘I’ as the grammatical subject, followed by the psy-

chological verb ‚believe‘ in the present tense indicative and by a syntactic complement, the embedded sentence ‘ $p$ ’. The second clause has the form of a ‘self-ascription’ of belief. ‘Self-ascriptions’ are first-person singular psychological ascriptions, i.e. ascriptions that exhibit ‘I’ as the grammatical subject of the main clause. The two clauses of (1) are connected by the conjunction, ‘but’, which is analyzed as a standard truth-functional logical connective. Thus, (1) has the schematic form:

(1) ‘ $p$  & I don’t believe that  $p$ ’.<sup>9</sup>

*Prima facie*, the two conjuncts are not logical contradictories, because they could both be simultaneously true. For example, it could be the case that it is raining,  $p$ , and yet I not believe that it is raining or believe that it is not raining. Moreover, others could comment on this situation, e.g. “It’s raining, but she doesn’t believe it”. Clearly, it is one thing for  $p$  to be true and quite another for me to *believe* that  $p$  is true. Accordingly, the two conjuncts of (1) are truth-functionally independent of one another on this construal:

$p$	‘I don’t believe that $p$ ’
F	T
T	F
F	F
T	T

If this is correct, then on “Frege’s criterium” for sameness of content, i.e. sameness of content entails sameness of truth value, ‘ $p$ ’ and ‘I believe that  $p$ ’ have different contents. On this analysis, the two conjuncts are logically consistent. Hence, the oddity of (1) cannot be traced to its expressing a formal contradiction.

This account assumes that the two conjuncts are about logically independent subject matters, i.e., ‘ $p$ ’ says something about the world, and ‘I don’t believe that  $p$ ’ says something about me. However, one could argue that, contrary to first impressions, both conjuncts have the *same* subject matter and that, consequently, (1) is really a contradiction in disguise; its surface structure is misleading. Consequently, any tokening of (1) in speech or in thought would be self-contradictory. Taking a hint from a remark of Wittgenstein’s,<sup>10</sup> one could deny that ‘I believe that  $p$ ’ is really a statement about the speaker and claim that the prefacing phrase ‘I believe that’ is to be understood, instead, as expressing the speaker’s insecurity or reservation as to whether  $p$ . On this view, proposed by Arthur Collins, the phrase ‘I believe that’ qualifies the  $p$ , which follows it.<sup>11</sup> It is claimed that, properly understood, (1) does have the form of a contradiction, ‘ $p$  & not- $p$ ’, because the phrase ‘I believe that’ has no semantic role. Hence, when someone tokens (1), her utterance or thought *is* self-contradictory. Moorean sentence (1) is

thus necessarily false and the air of paradox is dispelled. Is this analysis satisfactory?

Four considerations show that it is not. First, although there is a use of ‘I believe that’ to express reservation with respect to ‘*p*’ as well as a use which connotes deference in correcting someone else (“Actually, I believe that he came in second”), these are not the only uses of the phrase. In these uses the word ‘believe’ is stressed. When the word ‘I’ is emphasized (“**I** believe that *p*”) rather than the word ‘believe’, the use of ‘I believe that’ does not express reservation or uncertainty—quite the contrary. The latter use is not plausibly construed by claiming that ‘I believe that’ attenuates the ‘*p*’ which follows. Moreover, the fact that paralinguistic features like stress and intonation play a crucial role here suggests that a satisfactory diagnosis of the Paradox must go beyond the semantic features of Moorean sentences to consider matters of pragmatics.

Second, on the proposed analysis ‘believe’ does not function as a psychological verb, but as a modifier prefacing *p*. As Richard Moran points out, this has the consequence of barring me from talking or thinking about my own psychological condition, although I am able to talk and think about that of others.<sup>12</sup> This is certainly a counterintuitive consequence.

Third, the analysis makes the verb ‘to believe’ equivocal. First-person and third-person present tense uses of ‘believe’ differ in meaning. On this analysis, in first-person ascriptions the verb ‘believe’ does not concern anything psychological, whereas it *does* in other grammatical persons, as well as in first-person singular ascriptions in other tenses. Giving up the univocity of the belief predicate seems an unwarrantedly high price to pay for this particular solution to the Paradox. There are also important continuities between first-person and third-person ascriptions of belief. For example, the parallels between first- and third-person belief ascriptions make such a radical asymmetry improbable.<sup>13</sup>

Fourth, belief ascriptions in grammatical persons other than the first-person singular appear to have descriptive content and thus be truth-evaluable, e.g. ‘She believes that *p*’ is generally taken to describe something about a person and to be true or false, unless we deny the reality of the mental all together. Your ascription to me of ‘She doesn’t believe it is raining’ and my first-person ascription ‘I don’t believe that it is raining’ would seem to be true or false in the same circumstances. Due to its similar vocabulary and grammatical form, one would expect my first-person present tense ascription of belief to be descriptive and truth-evaluable, as well, just as first-person ascriptions in other tenses appear to be.

Accounts that try to explain Moore’s Paradox at the semantic level by reinterpreting the linguistic phrase ‘I believe that’ as a “preface” are unsatisfactory. The oddity of Moorean sentence (1) does not seem to be due to its semantic features. The objections to the “prefacing” account suggest that the oddness of uttering or thinking (1) has to do with the tokening of (1).

Moore himself describes what is “paradoxical” in terms of assertion. Adapting a remark of Wittgenstein’s, it might be claimed that the Paradox is due to the fact that ‘I believe that  $p$ ’ “says roughly the same as  $\neg p$ ”.<sup>14</sup> The idea is that there is an internal or necessary conceptual connection between assertion and belief. This is a view shared by many; Michael Dummett articulates a common assumption when he claims, “Assertion is rightly called an expression of belief”<sup>15</sup> If there is a necessary conceptual connection between assertion and belief, then saying ‘I believe that  $p$ ’ is equivalent to asserting ‘ $p$ ’ and saying ‘I don’t believe that  $p$ ’ is equivalent to asserting ‘ $\neg p$ ’. Uttering (1) is consequently equivalent to asserting ‘*I believe that  $p$  & I believe that  $\neg p$* ’, because in the first-person case asserting that  $p$  comes to the same thing as believing that  $p$ . Thus, what is asserted by utterances of Moorean sentences is self-contradictory. There are at least two reasons to resist this broad semantic account.

The first has to do with the sentences used and what they express. Although the sentences ‘I believe that  $p$ ’ and ‘ $p$ ’ may have similar *uses* in a given context, the sentences themselves have different meanings and different truth-conditions; for it seems clear that  $p$  can be true, even though I don’t believe  $p$  is true, and vice versa. Moreover, it is unclear how one is to get from the sentence ‘ $p$ ’, which is not tied to any point of view, first-person or otherwise, to the first-person or to something about belief on the basis of the sentence ‘ $p$ ’ alone. The alleged connection concerns at most pragmatic features of Moorean sentences, assertion and self-ascription, not their semantics.

Second, the crux of this account is the claim that there is a necessary conceptual connection between belief and assertion such that asserting that  $p$  entails believing that  $p$ . Surprisingly little argument is supplied for this claim. For it must be shown that asserting that  $p$  is sufficient for believing that  $p$ , because a speaker can dissemble, lie, or otherwise conceal his beliefs regarding  $p$ , asserting ‘ $p$ ’ all the while. Although people are usually prepared to assert what they believe, they do not always assert what they believe to be true, and sometimes they knowingly assert what they believe to be false. Therefore, a speaker’s assertion that  $p$  is not necessarily the expression of her belief that  $p$ . What is more likely is that we as hearers take unmarked utterances of indicative sentences to be assertions and take an assertion to be an expression of the speaker beliefs for inductive reasons, much as Moore suggested. Consequently, in asserting or stating that  $p$  I can get others to take me to believe that  $p$ , even if I don’t believe  $p$  myself. In sum, the connection between belief and assertion seems to be much weaker than this account allows.

Without additional argument for the claim that there is an internal connection between first-person belief ascription and assertion, this account does not provide a satisfactory explanation of the Paradox. However, it does suggest that an explanation of the Paradox must take features of the token-

ing of Moorean sentences or thoughts into account. Can the oddness of (1) be explained in terms of the pragmatic conditions for the speech act of assertion?

Many think so and construe Moore's Paradox as the paradigm of a pragmatic paradox. Such pragmatic construals explain the Paradox in terms of the self-defeat of the purposes of assertion. John Searle explicitly advances this view in *Speech Acts*, where he attributes the oddness of (1) to the speaker's violating the sincerity condition for the speech act of 'assertion'. On his view, whenever there is a psychological state specified in the sincerity condition of a speech act—believing to be true in the case of assertion, the performance of the act counts as an expression of that psychological state. Thus, to assert that  $p$  makes it the case that  $p$  counts as an expression of the belief that  $p$ . According to Searle, this "law" explains why I cannot assert both  $p$  and that I do not believe that  $p$ , thereby providing the solution to Moore's Paradox.<sup>16</sup> On this view, honest or sincere assertion requires actually having the belief expressed. However, this requirement on assertion seems too strong, because it is questionable whether sincerity really is a necessary condition for the speech act of assertion.

In the accounts of assertion inspired by Grice's analysis of non-natural meaning, a speaker asserts that  $p$  if and only if she produces an utterance with the intention to produce or activate a certain response in her addressee, e.g. a belief, on the basis of the addressee's recognition of the speaker's (complex) intention to produce the response.<sup>17</sup> Importantly, the addressee's recognition of the speaker's intention must function as part of the reason for addressee's producing the response. Assertion is thus analyzed in terms of communication between a speaker and an addressee. On the Gricean analysis, the oddness of an assertion of (1) is due to the fact that there is a conflict between the beliefs that the speaker must intend the addressee to recognize in order to take the utterance to be an assertion. For the second clause 'I don't believe that  $p$ ' seems to cancel the belief expressed by uttering the first clause ' $p$ '. Uttering a Moorean sentence as a means of getting the addressee to believe something through recognition of the speaker's belief and her complex intention that the addressee believe that  $p$  on the basis of the utterance produced is consequently self-defeating. Moreover, uttering a Moorean sentence violates a rationality constraint of the original Gricean account, namely, that the speaker must have reason to believe that she can fulfill her intention by producing the utterance in question.

On both of these pragmatic accounts, the assertion of (1) is pointless or self-defeating. Crucially, both pragmatic explanations of the Paradox require that an effect be produced in a hearer or addressee in order for an assertion to be made. Thus, they only apply to overt utterances in communication. This is as true of Searle's account as of explicitly Gricean accounts, for Searle builds the Gricean mechanism into his analysis of speech acts with only slight modification.<sup>18</sup> Neither account applies to thought. If Moorean

sentences have an analogue in thought, the speech act and Gricean explanations of the Paradox are incomplete at best. As they stand, they do not provide an adequate explanation.<sup>19</sup> *Mutatis mutandis* for accounts of the Paradox in terms of conventional implicature and conversational implicature.<sup>20</sup>

Although the speech act and Gricean accounts do not succeed in providing a satisfactory explanation of Moore's Paradox in speech and thought, they focus attention on the *tokening* of Moorean sentences. This is the right place to start. Both clauses of a Moorean sentence are tokened in an utterance or thought event. In tokening (1), the same person produces both '*p*' and 'I do not believe that *p*' in one utterance or thought. From the perspective of a hearer – a third-person point of view, '*p*' and 'She does not believe that *p*' can be simultaneously true, because '*p*' is a claim about the way things stand independently of a person's point of view and a person may well not believe what is the case; she may be in error. However, from a first-person point of view the belief that *p* at a particular time simply is the person's conviction about what is the case, about how things are. She can't get outside her own point of view to pronounce on how things are independently of her point of view.

When someone says 'I believe that *p*', she endorses *p* as the way things are. This is reflected in what has been called the "transparency" of belief to the world. As Gordon, following Wittgenstein and Gareth Evans, has emphasized, when someone asks me, "Do you believe that *p*?", I consider whether *p* is the case and answer "yes" or "no" accordingly. What has not been emphasized is that *I* have no alternative to my view that *p*. In the first person 'I believe that *p*' and '*p*' coincide in that they can be correctly tokened in the same circumstances. This is *not* because asserting that *p* is equivalent to believing that *p*, but rather due to the fact that in the first-person no distinction can be made between what I take to be the case and what actually is the case independently of what I may think, on a single occasion of tokening. Both '*p*' and 'I believe that *p*' express the first-person point of view when tokened by an agent, even though we tend to construe '*p*' from a third-person point of view, because it is unmarked.

When I express my view of how things are by tokening 'I believe that *p*', I could equally express my view of how things are by tokening '*p*'. Thus, when I token (1) "*p*, but I don't believe that *p*", I produce an utterance or thought which is nonsense, or "absurd" in the words of G. E. Moore. For in tokening (1) I cannot step outside of my view of how things are to pronounce on how they are independent of the way I take them to be. There is no possibility of comparison, no possibility of believing truly *vs.* believing falsely *within* the perspective expressed in a single tokening of a first-person present indicative utterance or thought. All I have to go on is what I take to be the case. I think this is what Wittgenstein has in mind when he says, "If there were a verb meaning 'to believe falsely', it would not have any significant first-person present indicative".<sup>21</sup> This is not to deny that the notion

of error or fallibility is central to the notion of belief, but to recognize the special situation involved in tokening first-person present tense expressions of (occurrent) belief. It suggests that the key to the oddity of Moore's Paradox lies in the first-person point of view and its role in the expression of belief.

## Notes

- <sup>1</sup> K. Linville and M. Ring 1991, "Moore's Paradox Revisited" *Synthese* 87, 273-83; Jane Heal 1994, "Moore's Paradox: A Wittgensteinian Approach", *Mind* 103, 5-24; Richard Moran 1997, "Self-Knowledge: Discovery, Resolution, and Undoing", *European Journal of Philosophy* 5, 141-161.
- <sup>2</sup> Jane Heal 1994.
- <sup>3</sup> Robert Gordon 2000, "Sellar's Ryleans Revisited" in: *Protosociology* 14, 102-114.
- <sup>4</sup> G. E. Moore 1942, "Reply to My Critics" in: P. Schilpp (ed), *The Philosophy of G.E. Moore*, LaSalle, IL: Open Court, 540-3; "Russell's Theory of Descriptions" in: P. Schilpp (ed.), *The Philosophy of Bertrand Russell*, LaSalle, IL: Open Court, 204.
- <sup>5</sup> G. E. Moore 1942, "Reply to My Critics", 542.
- <sup>6</sup> G.E. Moore, 1942, 542-3.
- <sup>7</sup> G. E. Moore 1942, 204: "To say such a thing as 'I believe he has gone out, but he has not' is absurd. This, though absurd, is not self-contradictory; for it may quite well be true. But it is absurd, because, by saying 'he has not gone out' we *imply* that we do *not* believe that he has gone out, though we neither assert this, nor does it follow from anything we do assert."
- <sup>8</sup> Heal 1994, Linville and Ring 1991.
- <sup>9</sup> Moorean sentence (2) has the schematic form ,I believe that  $p$  & not- $p$ ', respectively, ,not- $p$  & I believe that  $p$ ' and thus differs from (1). Although  $p$  is the contradictory of not- $p$ ,  $p$  falls within the scope of the propositional attitude operator 'believe', while 'not- $p$ ' stands alone as a conjunct. Consequently, the second version of the paradox also does not seem due to the two clauses being logical contradictories; for, like the first version, the two clauses are logically consistent.
- <sup>10</sup> Ludwig Wittgenstein writes, "One may say the following queer thing: 'I believe it is going to rain' means something like 'It is going to rain'. ...Moore's paradox may be expressed like *this*: 'I believe  $p$ ' says roughly the same as ' $\neg p$ ' but 'Suppose I believe that  $p$ ' does not say roughly the same as 'Suppose  $p$ ,'" in: *Remarks on the Philosophy of Psychology*, Oxford: Blackwell 1980 vol. I, 91-92.
- <sup>11</sup> Arthur W. Collins 1987, *The Nature of Mental Things*, Notre Dame, IN: University of Notre Dame Press, chap. 2.
- <sup>12</sup> Richard Moran 1997: 145.
- <sup>13</sup> Cf. L. Röska-Hardy, "Idealism and the 'I' of Self-Ascription", in: A. Øfsti, P. Ulrich, T. Wyller (ed.), *Indexicality and Idealism*, Paderborn: mentis 2000, 53 – 68.

- <sup>14</sup> Ludwig Wittgenstein, *Philosophical Investigations* Part II.x: 190; *Remarks of the Philosophy of Psychology* vol. I, 91.
- <sup>15</sup> M. Dummett 1973, *Frege's Philosophy of Language*, London: Duckworth, 330; B. Williams 1973, *Problems of the Self*, Cambridge: Cambridge University Press, 137.
- <sup>16</sup> J. R. Searle 1969, *Speech Acts*, 64-65.
- <sup>17</sup> Stephen Schiffer 1972, *Meaning*, 9, 39.
- <sup>18</sup> Cf. J. R. Searle 1969, 60, chap. II; Schiffer 1972, chap. IV.2.
- <sup>19</sup> Thomas Baldwin 1990, *G. E. Moore*, London/New York: Routledge, 2226-232, tries to extend the Gricean approach to thought via rationality considerations.
- <sup>20</sup> H. P. Grice 1975, „Logic and Conversation“, *Syntax and Semantics: Speech Acts*, 41-58.
- <sup>21</sup> L. Wittgenstein 1956, *Philosophical Investigations*, part II.x, 190.

## References

- Collins, Arthur W., 1987, *The Nature of Mental Things*, Notre Dame, IN: University of Notre Dame Press, chap. 2.
- Dummett, M., 1973, *Frege's Philosophy of Language*, London: Duckworth,
- Gordon, Robert, 2000, "Sellar's Ryleans Revisited" in: *Protosociology* 14, 102-114.
- Grice, H. P., 1975, „Logic and Conversation“, in: *Syntax and Semantics: Speech Acts*, 41-58.
- Heal, Jane, 1994, "Moore's Paradox: A Wittgensteinian Approach", *Mind* 103, 5-24
- LaSalle, IL, "Russell's Theory of Descriptions" in: P. Schilpp (ed.), *The Philosophy of Bertrand Russell*, LaSalle, IL Open Court.
- Linville, K. and Ring, M. 1991, "Moore's Paradox Revisited" *Synthese* 87, 273-83
- Moore, G. E., 1942, "Reply to My Critics" in: P. Schilpp (ed), *The Philosophy of G.E. Moore*
- Moran, Richard, 1997, "Self-Knowledge: Discovery, Resolution, and Undoing", *European Journal of Philosophy* 5, 141-161.
- Röska-Hardy, L., "Idealism and the 'I' of Self-Ascription", in: A. Øfsti, P. Ulrich, T. Wyller (ed.), *Indexicality and Idealism*, Paderborn: mentis 2000, 53 – 68.
- Williams, B. 1973, *Problems of the Self*, Cambridge: Cambridge University Press
- Wittgenstein, Ludwig, 1980, *Remarks on the Philosophy of Psychology*, Oxford: Blackwell 1980 vol. I, 91-92.
- Wittgenstein, L., 1956, *Philosophical Investigations*.