Trinity University

Digital Commons @ Trinity

Believing the Impossible

Curtis Brown
Trinity University, cbrown@trinity.edu

Repository Citation

BELIEVING THE IMPOSSIBLE

Curtis Brown

Ruth Barcan Marcus has argued that, just as we cannot know what is false, we cannot believe what is impossible.1 I will offer an interpretation of her defense of this view. I will then argue, first, that if the defense succeeded it would also justify rejecting many, perhaps most, of our ordinary belief ascriptions; and second, that, luckily, the defense does not succeed. Finally, I suggest that despite its failure there is something correct and important in Marcus's argument.

A prefatory note: I will speak of the objects of belief as "propositions." Marcus disapproves of this terminology, taking it to suggest that the objects of belief are linguistic or quasi-linguistic. Her preferred view is that the objects of belief are *states of affairs*, where these might be considered to be something like sequences of actual objects and properties. But I intend to use the term 'proposition' not for sentences but for whatever semantic entities are expressed by sentences on particular occasions of use; on this usage, states of affairs provide one among various possible explications of the nature of propositions. This use of the term "proposition" is fairly widespread, especially in recent writing (as Marcus herself notes: see 325, note 5).

To consider Marcus's defense, we need to remind ourselves of Kripke's well-known puzzle about Pierre's beliefs. Pierre, a native French speaker, assents, on good grounds, to the French sentence 'Londres est jolie'. Later he moves to England, learns English directly, without translation into French, and comes to assent to the English sentence 'London is not pretty'. He does not know that 'Londres' and 'London' refer to the same city. It seems that we have good reason to attribute to him both the belief that London is pretty (because of his assent to the French sentence) and the belief that London is not pretty (because of his assent to the English sentence).

In support of her view that we cannot believe the impossible, Marcus defends the intuition that one should give up the claim to believe something on learning that it could not possibly be true. For example, Pierre might well *say* "London is different from Londres." But, Marcus says, "in cases like Pierre's, once an impossibility had been *disclosed*, I would say that I had only *claimed* to believe that London was different from Londres. For that

would be tantamount to believing that something was not the same as itself, and surely I could never believe *that*" (330).2

In the course of defending her intuition, Marcus makes some important points. In particular, she argues, first, that linguistic acts like assent and dissent are only some among many indicators of belief, completely absent in some believers (dogs, for example) and not decisive even for those of us who do have language. As a result there is a difference between assenting to a sentence and believing the proposition it expresses. Second, she argues against a narrowly linguistic conception of rationality, one which pertains only to one's facility at manipulating sentences according to logical rules. Instead, "*we*will say that an agent is perfectly rational if *all* the behavioral indicators of belief are 'coherent' with one another" (334).

I believe both these points to be correct and important. They provide common ground by reference to which we may evaluate the considerations she marshalls in favor of her intuition that we cannot believe the impossible. How does Marcus defend this intuition? I will quote a crucial passage and then explain what argument I take the passage to be suggesting. Marcus writes, of the supposition that one can only believe what is possible:

> How otherwise is one to distinguish Pierre's *other* behavior, such as his "choice" behavior, from that of a wholly irrational agent? A *wholly* irrational agent is not merely *logically* irrational . . . he is an agent for whom linguistic and nonlinguistic belief indicators are incoherent. There is a question here whether, with widely incoherent belief indicators, he has beliefs at all. For what work could "belief" be employed in such a case? . . . [Marcus asks us to assume that Pierre is rational in the broad sense. Even so, she says] some of his behavior, and some of the intended outcomes of his purported beliefs about London, would be just like those of an irrational agent. If, for example, Pierre, in accordance with his assents, were to bet on a given occasion onLondres being pretty and London being not pretty, the *outcome* of his bet would be as self-vitiating as that of the wholly irrational agent. Yet purely *logical* considerations and the empirical justification available to him made that bet wholly "reasonable" (336-337).

Let us begin to explicate this by elaborating a bit on Marcus's wager example. We imagine a psychological experiment with two subjects, Pierre and his irrational counterpart Ira. Ira's history is much like Pierre's (born in France, moved to England, learned English without translation, and so on), but Ira,

unlike Pierre, believes that 'Londres' and 'London' are coreferential. Each of the subjects is given a sheet of paper containing four sentences, and asked to check one of them. Each is told that he will be paid $10 if the sentence he checks is true and charged $10 otherwise. The sentences are:

Londres is pretty and London is not pretty.

Londres is not pretty and London is pretty.

Londres is not pretty and London is not pretty.

Londres is pretty and London is pretty.

Confronted with the sheet of paper, let us suppose, Pierre and Ira both check the first sentence. Pierre checks it because he assents to the sentence, and he assents to the sentence for the Kripkean reasons already discussed. Why does Ira check the first sentence? He believes that 'Londres' and 'London' are coreferential, and so realizes that the sentence is contradictory: he either accepts the contradictory sentence anyway, or else bets on the truth of a sentence he does not accept. Either way his behavior seems clearly irrational. I take Marcus to be asserting that we can correctly distinguish Pierre's behavior and the reasons for it from Ira's genuinely irrational behavior only by rejecting the idea that Pierre believes the impossible.

If this is indeed Marcus's strategy, then she needs to show two things: first, that if we reject the idea that Pierre believes the impossible, we can explain his actions (but not Ira's) in a way that reveals their rationality; second, that if we do *not* reject the idea that Pierre believes the impossible, we *cannot* explain his actions in a way that reveals their rationality, that is, in a way that distinguishes them from Ira's. Consider the first point: suppose that Pierre does not believe anything impossible, although he assents to a sentence which, unknown to him, expresses an impossible proposition. Then we can explain his behavior in a way which exhibits its rationality: he wants to receive $10, he believes that if he places a check mark by a true sentence he will receive $10, and he believes that of the sentences before him the first and only the first is true. So, quite reasonably, he engages in the behavior of placing a check by the first sentence. No such explanation is available for Ira, who knows the sentence is contradictory, knows it cannot possibly be true, but checks it anyway, either because he believes things he knows cannot be true or because he gambles on the truth of statements he believes to be false.

So far our results support what I am taking to be Marcus's argument. If we do not assume Pierre believes the impossible, we can explain why his behavior is rational while Ira's identical behavior is not. Now consider the second point: suppose that Pierre *does* believe the impossible, that he *does* believe that Londres is pretty and London is not pretty. It seems that there is no explanation of the rationality of Pierre's action which appeals to his believing that Londres is pretty and London is not pretty. If we try to explain his bet by saying "he bet on the impossible proposition because he believed it," this makes him sound irrational--quite as irrational, in fact, as Ira. This, I take it, is why Marcus claims that if we attribute to Pierre a belief in the impossible we will not be able to distinguish his behavior "from that of a wholly irrational agent." (But how does *adding* the impossible proposition in question to the things we take Pierre to believe invalidate the explanation I have already offered of the difference between Pierre and Ira? Pierre still has the beliefs that explanation made use of, and they still rationalize his behavior. I will return to this point.)

My response to this argument is twofold. I will argue, first, that the very same sort of argument would justify denying far more of Pierre's beliefs than just his beliefs in the impossible. In fact it would justify rejecting many, perhaps most, of our ordinary belief ascriptions. But I will also argue, second, that we *can* explain the difference between Pierre and Ira without denying that Pierre believes the impossible.

First: Marcus's argument, if successful, shows far more than she wants it to. While Marcus denies that Pierre believes the impossible, she accepts that he has *incompatible* beliefs: he believes thatLondres is pretty, that is, he believes that London is pretty; and he also believes that London is not pretty.3 And this is enough to produce the same sort of difficulties she finds unacceptable in the case of belief in the impossible. Suppose that instead of being asked to check one of the four sentences listed earlier, Pierre and Ira are asked to check as few or as many sentences as they like from the following list, and told that they will be paid for every truth they check and charged for every falsehood. The list consists of:

London is pretty.

    London is not pretty.

    Londres is pretty.

    Londres is not pretty.

Pierre and Ira both check the second and third sentences, thus guaranteeing that they will win nothing. Shall we say that Pierre checks the second box because he believes that London is not pretty, and checks the third box because he believes London is pretty? If so, how is this different from our explanation of Ira's choice? Surely any explanation of Pierre's behavior that accounts for its rationality will need to appeal to the fact that he thinks both his choices can be true together. Shall we say that Pierre believes these incompatible propositions to be compatible? That seems to exacerbate rather than eliminate his apparent irrationality. And if it *did* succeed in exhibiting the rationality of Pierre's actions, then Marcus could also exhibit his rationality in the first gambling case by saying that Pierre there believed an impossible proposition to be possible. Here, as before, if we want to exhibit the rationality of Pierre's behavior, we do best to retreat to his attitudes toward sentences, pointing out that he believes the second and third sentences are true and that he also believes these two sentences are compatible. It is precisely this inability to explain Pierre's reasonableness in terms of his beliefs about London that Marcus uses as justification for denying that he believes the impossible, so the case that Pierre also does not believe individually that London is pretty or not pretty seems equally strong.

There are also less artificial examples of Pierre's apparent irrationality. On Marcus's account, Pierre believes that London is pretty (in virtue of accepting 'Londres est jolie'). Why then does he refuse to assent to 'London is pretty', a sentence he understands and which expresses a proposition he believes? Why does he mutter "ugh" under his breath whenever he goes for a walk in London? These seem like precisely the sorts of apparent incoherence among belief indicators which Marcus hopes to avoid by rejecting belief in the impossible. And here once again, to explain why Pierre is rational while Ira is not, we will need to withdraw from the level of belief in the proposition that London is pretty to the level of beliefs about sentences like 'Londres est jolie'. Again, Marcus has as good a justification for rejecting Pierre's beliefs that London is pretty and that London is not pretty as she does for rejecting his belief that London is pretty *and* London is not pretty.

We have seen that Marcus's argument, if successful, would require us to deny not only that one can believe the impossible, but also, contrary to her intention, that one cannot have incompatible beliefs. But the skeptical consequences of Marcus's argument do not end there. If Marcus were to deny the possibility of incompatible beliefs, as I have argued she must, she would also have to deny the possibility of beliefs with which one *would* come to believe something incompatible.

Why does she need to claim this? Consider the point at which we would ordinarily say that Pierre acquired the belief that London is not pretty. If it's impossible to have incompatible beliefs, then we must deny *either* that he really believes that London is pretty, or that he believes that London is not pretty, or both. Since the credentials of these two beliefs are equally good, for instance, in the case of Pierre, we will need to deny that Pierre believes *either*. So either he suddenly loses the belief that London is pretty or he *never* had it. But it can't be that he suddenly lost it, since he has not undergone any of the changes by which one loses a belief, such as forgetting it or changing one's mind. So he must never have had it.

Now consider the time at which it *seemed* (falsely, as it turns out) that Pierre acquired the belief that London is pretty. *Either* the fact that he doesn't believe that London is pretty depends on something that hasn't happened yet, namely his later apparent acquisition of the belief that London is not pretty, or it does not. But surely what one now believes cannot depend on something that has not yet happened, and will not happen for months or years. If this is correct, then we need a principle which, depending only on facts up to now, will rule out an agent's every believing things with which belief the agent later adds are incompatible. But the only principle which is guaranteed to accomplish this is the principle that one can't have any belief with which one *could* later come to believe something incompatible. But this in turn will lead to a wholesale rejection of almost *any* beliefs in singular propositions and also most beliefs we have in virtue of believing to be true sentences containing proper names, kind terms, mathematical expressions, and so on.

Supposing I have understood Marcus's argument correctly, then, it has much more wide-ranging and sceptical consequences than she acknowledges or intends--consequences so wide-ranging andsceptical that we might regard them as a *reductio* of her argument that we cannot believe the impossible. But then what of the problem of distinguishing Pierre's behavior from Ira's? Is there after all no difference, or at least no way to explain the difference?

I maintain that precisely the same explanation of Pierre's behavior is possible whether or not we deny that he has incompatible beliefs, and whether or not we deny that he believes the impossible. We do not need to deny that Pierre believes that London is pretty to explain the rationality of, for example, his refusal to assent to 'London is pretty'. Rather, we need to keep track of the deeper beliefs in virtue of which he believes that London is pretty. Pierre believes that the sentence 'Londres est jolie', of which he is a competent user, is true. It is *in virtue of* believing this (together with various further beliefs, for example about the meaning of the sentence, and together with various facts

about his circumstances, such as the fact that 'Londres' refers to London) that he believes that London is pretty: if he did *not* believe that the French sentence were true, then our grounds for attributing to him the belief that London is pretty would disappear. To explain his rationality, we need to focus not on his beliefs about London, but on the belief in the truth of the French sentence which grounds one of his beliefs about London. But this does not require rejecting the belief about London.

For similar reasons, we do not need to deny that Pierre believes that Londres is pretty and London is not pretty in order to explain the difference between his behavior and Ira's on our first gambling experiment. To exhibit Pierre's rationality we need, not to deny his belief in the impossible, but to appeal to the *way* he believes it, to the deeper beliefs in virtue of which he believes it--in particular, in this case, the belief that 'Londres est jolie' is true *and* 'London is pretty' is true. This deeper belief *could* be true. So Pierre, unlike Ira, believes something impossible in virtue of believing something possible; his behavior can be seen to be reasonable in light of this deeper, possible, belief. Ordinarily we can make use of people's indirect beliefs in explaining their behavior without getting into difficulty, but when, as with Pierre, puzzles arise, we need to look deeper, at the less indirect beliefs in virtue of which we believe our more indirect ones.

There is a simple argument from two premisses to a conclusion which is useful in thinking about the issues Marcus raises. I suspect that something like this argument underlies her thinking. It goes like this:

> (1) The objects of one's belief characterize one's mental state, and

> (2) Impossibilities could not usefully characterize a rational person's state of mind; so

> (3) A rational person cannot believe the impossible.

(Marcus suggests a less qualified conclusion, but her case is strongest as applied to rational individuals.)

Consider the two premisses of our argument, beginning with (1). "Mental state" is to be taken to be one's *narrow* or *intrinsic* mental state, the sort of mental state that cannot change unless something inside the person changes. Conceivably this is a Cartesian chimera, but it is surely the ordinary notion of a mental state. And the claim that the objects of one's belief characterize one's mental state is to be taken in a very strong sense: the things

one believes could not be different unless one's mental state were different, and one's belief state could not be different unless the objects of one's belief were different.4

This is a strong interpretation of what it takes to characterize one's mental state, but it is the interpretation we need to make (2), to which we now turn, plausible. There are two reasons for the plausibility of the claim that impossibilities cannot usefully characterize a rational person's state of mind: one appeals to belief alone, the other to the relation between belief and action. The first reason is that the most basic rule concerning the rationality of belief is something like "don't believe what cannot possibly be true," so it is very hard indeed to see how an impossibility could characterize the mental state of a rational person.5 If objects of belief characterize one's mental state, we ought to be able to appeal to these objects of belief in exhibiting or explaining one's rationality, and for this purpose impossibilities simply will not serve.

The second reason for the plausibility of (2) concerns the relation between belief and action, and is suggested by Marcus's discussion of these matters. Belief is tied closely to action: we need to be able to explain one's actions in terms of one's beliefs and desires, and to make inferences about one's beliefs on the evidence provided by one's behavior. And the notion of rationality is deeply involved in both projects. If an individual is rational, then the beliefs that explain the individual's actions should also make them reasonable, and conversely the beliefs we may infer from one's actions are precisely those in terms of which the actions make most sense.

What, then, of impossibilities? What actions do they make reasonable? What behavior do they predict or explain? Take Pierre, for instance. He believes that London is pretty and that London is not pretty. What then will he do? Go to London? Try to stay away? Either action seems irrational given his beliefs, so neither can be predicted; if he performs either, his beliefs do not explain why. Each conjunct of the sentence I have used to describe his belief destroys the predictions or explanations that might have been made on the basis of the other conjunct,6 so that Pierre's belief as a whole seems to offer us no insight into his behavior. And it is just as unclear how one could infer belief in an impossibility from one's behavior. If one's behavior utterly stopped making sense, then perhaps we would have grounds for ascribing to that person belief in impossibilities if we had grounds for ascribing any beliefs at all. But Pierre's behavior is completely rational, so we need to ascribe beliefs according to which it makes sense; ascribing beliefs in the impossible will not accomplish this.

We have seen why the second premise in our little argument seems so plausible. Supposing it is correct, we have two choices, neither of which seems attractive. One possibility, the one Marcus adopts, is to reject the view that we can believe the impossible, that is, to accept (3). I have argued that the only grounds on which this can be justified will lead to a wholesale rejection of our ordinary ascriptions of belief. We can now see why from a slightly broader perspective. The motivation for rejecting belief in the impossible is that it fails to exhibit one's rationality. But almost any indirect belief at least runs the risk of this sort of failure. If we are to ascribe only those objects of belief essential to making sense of one's thoughts and actions, we will need to abandon far too much of our ordinary practice of belief ascription.

The second possibility is to reject premise (1), to reject the idea that objects of belief characterize our narrow mental states in the strong sense. This is the course suggested by much recent writing on belief.7 And it seems to me that this is the course we *must* take. Admitting the existence of indirect belief is tantamount to rejecting (1), and the evidence for indirect belief is overwhelming. But there are reasons for (1) that we need to find a way to preserve, and one of the virtues of Marcus's piece is to call our attention to these reasons. I suggest that while (1) is not true of belief in general, there is a subclass of our beliefs of which it is true, and of which, therefore, the thesis that we cannot believe the impossible is also true. This is the subclass consisting of our *direct* beliefs.

Why should we think there are direct beliefs, beliefs which characterize our mental state in the sense of (1)? We may offer two reasons parallel to the two reasons provided for (2). First, consider the rationality of belief. An account of the rationality of belief must be an account of the rationality of *what is believed*, of the objects of belief. If we talk about mental states in physicalistic terms, or even in syntactic terms, we simply lose our grip on the notion of rationality.8 Possibly we will ultimately be able to say that some patterns of neuron firings are irrational, but if so this will be only because we have correlated these patterns with certain belief contents, which we know on other grounds are irrational. And even at the syntactic level, if it makes sense to speak of sentences in the head, it will not make sense to criticize a subject for having, say, the sentences S and not S in the subject's "belief-box" *unless* we know that the syntactically identified mark "not" *means negation*. Talk about the rationality of belief seems to be irreducibly talk about the *contents* of belief, about what is believed.

We have seen that indirect beliefs always run the risk of concealing the reasonableness of our being in the belief states we are in, of making what is

rational appear irrational. Since we need to evaluate rationality in terms of the contents of belief, but cannot reliably do so with respect to indirect belief, we ought to conclude that rationality is best assessed in terms of our direct beliefs. And this gives us some reason to think there are direct beliefs.

A second, similar, reason concerns the rationality of action. We need to be able to evaluate actions as rational or irrational, and to do this we need explanations of action that appeal to the *contents* of our beliefs and desires. It may be possible to explain behavior in purely mechanical terms, but we cannot judge its reasonableness without appealing to objects of belief. We have seen how indirect belief always runs the risk of failing to reveal the reasonableness of our actions. If objects of belief are essential, but mediate objects of belief will not do, that provides some reason for supposing that there are immediate objects of belief. 9

NOTES

1. Ruth Barcan Marcus, "A Proposed Solution to a Puzzle About Belief," *Midwest Studies in Philosophy* 6 (1981): 501-510, and "Rationality and Believing the Impossible," *Journal of Philosophy* 80 (1983): 321-338. (Page references in the text will be to the latter.) For useful commentary on these papers see Ferenc Altrichter, "Belief and Possibility," *Journal of Philosophy* 82 (1985): 364-382; and ReinaldoElugardo, "Marcus's Puzzle About Belief-Attribution," *Southern Journal of Philosophy* 24 (1986): 201-218.

2. Nathan Salmon has argued that the first belief is *not* tantamount to the second: see his "Reflexivity," *Notre Dame Journal of Formal Logic* 27 (1986): reprinted in N. Salmon and S. Soames, ed.,*Propositions and Attitudes* (Oxford: Oxford University Press, 1988): pp. 240-274, at pp. 259-60.

3. This is fairly clear in "Rationality and Believing the Impossible" and quite explicit in "A Proposed Solution to a Puzzle About Belief."

4. I use the term 'belief state' in the second conjunct to allow that other facets of one's mental life--one's emotions, perhaps, or one's desires--might be different without the objects of one's beliefs being different.

5. Compare Richard Foley in "Is It Possible to Have Contradictory Beliefs?" *Midwest Studies in Philosophy* 10 (1986): 327-356, at 341: "One prerequisite of an adequate theory of belief is that in conjunction with a plausible theory of epistemic rationality it preclude the possibility of a person having beliefs that are both genuinely (i.e. irreducibly) contradictory

and epistemically rational." Cf. Foley, *The Theory of Epistemic Rationality* (Cambridge: Harvard University Press, 1987): at pp. 237-240. For discussion of Foley's views on believing the impossible, see my "How to Believe the Impossible," forthcoming in *Philosophical Studies*.

6. I do not mean to imply that the belief itself is conjunctive. On some accounts of the objects of belief, for example the possible-worlds account, the objects of belief will not have logical forms at all.

7. I have in mind the following: a series of articles by Tyler Burge, including "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979): 73-122; "Other Bodies," in Andrew Woodfield, ed., *Thought and Object* (Oxford: Oxford University Press, 1982): pp. 97-120; "Individualism and Psychology," *Philosophical Review* 95 (1986): 3-46; "Intellectual Norms and Foundations of Mind," *Journal of Philosophy* 83 (1986): 697-720; Lynne Rudder Baker, *Saving Belief: A Critique of Physicalism* (Princeton: Princeton University Press, 1987); and the essays collected in Philip Pettit and John McDowell, eds., *Subject, Thought, and Context* (Oxford: Oxford University Press, 1986).

8. Compare Daniel Dennett's view that rationality is relevant when we take up the "intentional stance" but not when we take up the "physical stance" or the "design stance." See e.g. Dennett, <u>The Intentional Stance</u> (Cambridge: MIT Press/Bradford Books, 1987).