We Are Reasons-Responsive Creatures: An Interview with Emma Borg

Bruno Mölder

Department of Philosophy, University of Tartu

Emma Borg has been a professor at the Institute of Philosophy, School of Advanced Studies, University of London since January 2024. Prior to that, she worked at the University of Reading starting in 1998. She completed her PhD at University College London. Borg's primary areas of research are philosophy of language, philosophy of mind, and business ethics. She is the author of *Minimal Semantics* (Oxford University Press, 2004), *Pursuing Meaning* (Oxford University Press, 2012), and *Acting for Reasons: In Defence of Common-Sense Psychology* (Oxford University Press, 2024). Emma Borg gave the Gottlob Frege Lectures in Theoretical Philosophy at the University of Tartu from November 4–6, 2024, under the title "Reasons for Action". The interview took place in Tartu on November 6, 2024.

Keywords: Borg, common-sense psychology, reasons-responsiveness

---

How did you first become interested in philosophy? Were there any particular experiences or events that led you to choose it as a profession?

Like many people who ended up as philosophers, I was probably one of those annoying children who liked to ask a lot of questions. Essentially, philosophers are the people who just keep asking questions and are never entirely convinced by the answers they get. There is one childhood experience that I remember very clearly. I remember being in the kitchen with my older brother and one of my parents. My brother said that a spaceship could appear in the room, and my mother replied, "That is ridiculous. It is impossible, it could never happen." My brother countered, "Nothing is impossible." I recall thinking, even at a young age, is it true that nothing is impossible? Or, if something really is impossible, what would that be like? I found that a

Corresponding author's address: Bruno Mölder, email: bruno.moelder@ut.ee.

© *All Copyright Author*

Online ISSN: 1736–5899
www.spe.ut.ee
https://doi.org/10.12697/spe.2025.18.01

fascinating question. That was probably the first instance of my philosophical thinking.

However, when I got to university, I was not planning on studying philosophy at all. I started off studying English, but I found it really disappointing—it did not fit with my intellectual interests at all. I only came to philosophy because it was a subject I had done a bit of at school, and I thought it might be something I could do. I was an undergraduate at King's College London, and I remember very clearly going to the philosophy department and saying, "I have done a term of English, but I do not really want to stick with it. Would philosophy have me?" They said yes, and as soon as I started, I discovered that philosophy was the subject I really liked.

Emma Borg in Tartu in 2024. Photo by Bruno Mölder.

Who were your teachers at King's?

At King's, one particularly influential teacher was Mark Sainsbury. He was the first person I studied philosophy of language with. I found the fundamental questions in philosophy of language really interesting: What is the meaning associated with fictional expressions? How do we understand empty names like "Hamlet" or "Pegasus"? What grounds the meaning of a linguistic expression? I think philosophy of language is one of the most difficult bits of philosophy because it is so meta. You have to express it in language, but it is about language. I found that combination of factors really interesting to think about.

As you said, you worked in the philosophy of language. What were your main areas of interest in that field, and how would you summarize your key views?

Philosophy of language is definitely what I am most known for, if I am known for anything. It was actually only when I got a job at the University of Reading that I got really interested in questions about what speakers convey when they use language, how communication works and what content people are tracking when they are engaged in conversation exchanges. The issues are probably easiest to think about with an example. Charles Travis, who is an absolute genius at thinking up these kinds of examples, presents a thought experiment[1] where we imagine that I am making tea and need some milk (because I am British and have milk in my tea). I look at you and ask, "Do you have any milk?" You reply, "Yeah, there is milk in the fridge." I go to the fridge, open it, and expect to find a bottle of milk inside. There is no bottle of milk—just a small puddle of milk on the bottom shelf. The question Travis poses is this: Have you spoken truthfully or falsely? Have you said something true or something false? It seems that, because I was expecting to find milk in a form appropriate for making tea, Travis suggests that, when you said there is milk in the fridge, you have said something false. That kind of position became known as a contextualist stance. Many people adopted this idea that the content that you literally express should be viewed as contextually enriched. So you do not just say there is milk in the fridge, you say that there is milk in the fridge suitable for tea.

I thought that this gets something right about the conversational exchange, but we should not overlook the valuable role played by what I called the minimal content. The minimal content, I argued, is the content you get just from the meanings of the words and the way they are put together in the sentence. When you said there was milk in the fridge, you produced a sentence with a literal meaning: "There is milk in the fridge." When I opened it and found a puddle of milk, what you literally expressed was true. What failed was a pragmatic consideration—namely, I wanted the milk for tea, and that is not what I got. So the position I have become best known for in philosophy of language is the defence of minimal semantics, the claim that there are literal meanings that are attached to sentences. Those meanings are truth-evaluable. That means that if you hold them up against the world, you can see whether the world satisfies them or not. You can talk about the way in which the world would be if that sentence were literally true. Minimal meanings may diverge significantly from the contents that we communicate

[1] Travis, Charles (1989). *The Uses of Sense: Wittgenstein's Philosophy of Language*. Oxford: Clarendon Press, pp. 18–19.

with one another, but still they play an important role (for instance, in assessing what content people are committed to and when).

What motivated your shift from the philosophy of language to the philosophy of mind?

When I was doing philosophy of language in graduate school, I mostly thought about demonstratives and indexicals—expressions like "that fish," "this bottle," "I," "you," and "today." These kinds of expressions seem to derive part of their meaning from the context in which they are uttered.

When I was thinking about minimalism, I encountered a worry about how to accommodate these kinds of expressions, because they seem to require access to the mental states of the speaker. For example, when you say "that bottle" and we are faced with a row of bottles, it seems like I need to know which one you intended to refer to in order to identify the object. This appears to pose a challenge for minimalism, so I started thinking more about demonstratives, indexicals, and the role of speaker intentions in understanding language. That line of thought pushed me to consider more generally what it means to access mental states. And that in turn pushed me back to ask questions about common-sense psychology and the extent to which we need to think of ourselves as reasons-responsive creatures.

Your Frege lectures at the University of Tartu were titled "Reasons for Action." Could you briefly summarize the central argument of these lectures?

The claim I really want to defend is that we are generally reasons-responsive—not always, and not always properly—but typically, adult humans are the kinds of creatures who do what they do for the reasons they have. Maybe infants do that as well, and maybe certain kinds of animals. This position is known in philosophy as common-sense psychology or folk psychology. For a long time, it was the orthodox view, but it has come under significant pressure from a particular kind of empirical attack. This alternative approach argues that when you get out of your armchair and examine how people actually make decisions or how we understand the actions of others, you find that people are not acting on the basis of reasons, nor are they understanding each other through the lens of beliefs, desires, and other mental states.

The evidence for this is supposed to be empirical. It involves experiments that suggest we often seem irrational, driven by various heuristics and biases. These studies also suggest that we understand each other not in terms of beliefs and desires but in a simpler, unconscious way that does not involve

these states. What I try to do in the book and in the lectures is to show that we can resist these experimental arguments.

Why do you believe it is important to defend common-sense psychology? Why is it important to see ourselves as rational beings responsive to reasons?

There are two reasons I would give for that. One is that this reasons-based framework is extraordinarily practically successful. Jerry Fodor, whose philosophy I am generally a big fan of, points out that the framework of folk psychology, or common-sense psychology, allows us to move from very impoverished evidence to really rich social coordination.

The example[2] he gives is the following: someone rings me up and asks, "Do you want to lecture at my university?" and I say, "Yes, I do. I will be there on the 3 p.m. flight next Thursday." That is all that happens, and yet both of us will turn up at the airport at 3 p.m. on that Thursday—you to pick me up, and me to meet you. You have managed to predict my actions, I have managed to predict yours, and the social coordination has worked perfectly. As Fodor points out, if it does not work perfectly—if one or the other of us is not there—it is much more likely that something has gone wrong with the airline or the traffic rather than with the prediction folk psychology has made.

One thing that is special about philosophy is that it takes facts that, at first glance, seem totally unremarkable and uninteresting—like our ability to predict what someone else will do—and asks us to reflect on them. Once we do, we can see that this ability is amazing. Our capacity to interact with one another in such complex social settings is phenomenal. We should be impressed by it, and we should wonder how it happens. I think common-sense psychology works really well as an explanation of this amazing ability.

In addition, the idea that we are reasons-responsive creatures is absolutely fundamental to a range of core philosophical notions. It is fundamental to our conception of ourselves as agents—what it is to be an agent is to be able to respond appropriately to reasons. It is fundamental to our notion of what it is to be a person—a person is fundamentally a rational thinking being, a rational animal. It is also central to notions that matter in society and moral philosophy. It is hard to imagine how we could have concepts like praise, blame, and responsibility without thinking of ourselves as reasons-responsive and rational.

[2] Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in The Philosophy of Mind*. Cambridge, MA: The MIT Press, p. 3.

Is your claim that these views, influenced by cognitive psychology, are unable to take concepts like responsibility and blame into account? They could argue that one is blameworthy because one used the wrong heuristics this time, for instance.

One of the things I argue is that we need to get really clear on what the challenge from these alternative positions is. One argument is that heuristics are inherently irrational because they are not responsive to evidence. I think that is not quite the right argument to have. There are different ways we can conceive of cognitive heuristics. One way is to think of them as habits that do not appeal to reasons at all. If we understand them that way, they do turn out to be genuinely irrational because they are simply not in the space of reasons. That would indeed be a problem for our notion of responsibility.

But there are alternative views of what heuristics are—namely, that they are non-logical rules of thumb that allow us to make decisions with limited consultation of evidence. What I try to say about this way of understanding heuristics is that it may actually turn out to be a perfectly rational, reasons-responsive process because it does look at evidence. It might just be looking at limited evidence, and, on certain occasions, that kind of limitation may be fine. There is a distinction between ideal rational choice—where we aim to maximize results and always find the correct solution—and a notion of bounded rationality, which suggests we may only need to consider some evidence and make a good-enough judgment based on it. If bounded rationality is what is required for rationality, then heuristics might fit perfectly well within an account of what it is to be a rational creature. In that case, an individual might well be praiseworthy for using a good or appropriate heuristic and blameworthy for using a poor one, but note that that is only because heuristics still look to evidence. That is to say, our use of them is still reasons-responsive.

Some of the sciences of the mind, like theoretical psychology, also deal with similar questions. How do you view the relationship between the philosophy of mind and the sciences of the mind? What, if anything, sets philosophy apart?

The kind of philosophy I do is definitely at the border of many different disciplines. When I was doing philosophy of language, it bordered up with linguistics. My philosophy of mind research borders psychology—comparative psychology, social psychology, developmental psychology. The sort of philosophy I like pays attention to discoveries in other disciplines. I am not so interested in pure philosophy that does not engage with what people in other fields are saying.

In some ways, there is nothing uniquely special about the discipline of philosophy. Some of the people I read and find interesting are in psychology departments, some are in philosophy departments, and some are in other departments. In that sense, the divide is somewhat artificial, and we should not worry about it too much.

But on the other hand, I do think there is something special about philosophy—not necessarily as an academic discipline, but as the kind of thing that it is. Philosophers are concerned with asking really fundamental questions and with how an entire theory hangs together, in a way that you do not always find in psychology. For example, in the heuristics and biases literature, many experiments are fascinating and fun, but there is often a quick move from experimental results to claims like "we are not properly reasons-driven animals" or "we are irrational a lot of the time." If philosophy has a role here, it is in asking how we arrived at those conclusions from the evidence. We need to be quite careful about that process and think through exactly what claims we are committing to. So, I think there is a place for philosophers because they ask the big questions. Even if they cannot always answer them, they frame them well, slow down other disciplines from moving too quickly, and remind everyone that we are only human and there is a danger of making overblown claims.

Do you believe that other scientists are really listening to philosophers? Do they appreciate philosophers telling them how to do things?

This is an interesting question, and there might not be a general answer. I think it depends on the personality of the individuals involved, but I like to think at least some scientists are listening. When I was at the University of Reading, I used to run a joint centre between philosophy and psychology, and those philosophers and psychologists certainly wanted to talk to one another. In many areas, there seems to be an appetite for learning from other disciplines.

Do you have a particular method or approach to doing philosophy? You mentioned interdisciplinarity.

Yes, interdisciplinarity really matters to me. I like philosophy that makes contact with empirical work. When I hear a paper in pure metaphysics, it often strikes me as the purest kind of philosophy because it is about conceptual analysis. Those questions still strike me as fascinating, but I do not think that is the kind of philosophy I am particularly good at. The kind of philosophy I like looks at experimental findings and asks: What do these findings

tell us about the human mind? What do they tell us about language, thought, and action?

Is there a place for intuitions in philosophy?

In some ways, yes. We talked a bit in the lectures about the notion of motivated reasoning—the idea that we use the beliefs we already hold to dictate our search for evidence and shape the way we respond to that evidence. Motivated reasoning is often seen as a bad thing, but I argued in the lectures that it is not always so. What we already believe often serves as a useful guide for directing our investigations. If you think of intuitions as starting hunches, then it seems right that they should help set the starting point.

What I like about philosophers is how incredibly self-critical they are. They want to defend a position, but their first step is to ask: What is wrong with it? I think that is a nice feature. Maybe that is the role intuitions play—they give us the starting point from which we apply our critical faculties.

The intuitions of ordinary people are also studied by experimental philosophy. How do you see that?

I have done some XPhi work. For instance, I have done questionnaires on pain and have been involved in XPhi epistemology projects with others. I feel slightly conflicted about it. On the one hand, I think it is great that philosophers are getting better at designing and conducting experiments. The slight flip side of wanting to do empirically engaged philosophy is that I think we should not be completely driven by lay responses. The aim of philosophy is to get at the truth, and that truth may not be just what laypeople think. So, there may be a gap between the two, but there is definitely a place for experimental work. If we want to know about the concepts people have, we should go and ask them questions.

Do you think philosophy has changed since the time you studied it? If so, how?

One could say that it has become more empirically based and interdisciplinary, but I am not sure that is entirely true. I think philosophers have always been quite good at engaging with other disciplines and drawing understanding from different areas.

I was thinking about larger philosophical trends, such as how, in the 1940s and 1950s, ordinary language philosophy was dominant, and later, the philosophy of mind gained prominence, especially with the rise of the computer metaphor. Have you noticed any similarly significant movements or trends that have emerged in the past 30 years?

I do not think there is anything as codified as the linguistic turn that has happened in my time as a philosopher. However, there has been a resurgence of philosophers being slightly more generalist. There was a period when philosophers became incredibly narrowly focused. Formal epistemology is a case in point. The idea was that you did not want to say anything wrong, so you made the smallest possible move you thought you could defend. That led to increasing epicycles of technicality and a focus on tiny issues within a group of like-minded theorists.

That approach is very different from someone like Hume. Hume had an incredible breadth of philosophy and a whole worldview about how everything hung together. While we may not be quite back at Hume, I think there is now more room for philosophers who want to address broader issues, rather than just dealing with very small, technical problems.

Another way philosophy has changed is its openness to more practical questions. For example, when I started, Davidson was all the rage in the philosophy of language, and the focus was on writing T-sentences and doing very formal work. Now, philosophers of language are much more interested in questions like the nature of slurs, what happens to content when language is used online, and how to understand gendered expressions. So, there has been a shift in philosophy toward being open to these more applied questions.

Do you believe there is progress in philosophy? If so, in what ways?

What I like about philosophy is that progress is not really measured by how many questions we have conclusively answered, which is lucky because we probably have not conclusively answered all that many questions. I think the way progress happens in philosophy is that someone makes a shift, notices something, or comes up with an account, and for a while, there is a massive amount of energy and activity around that question or theory. Then, that question does not get resolved, but people just get a bit bored with talking about it and move on to something else instead.

However, if you hang around long enough, what you find is that after a while, that area and those questions become interesting again. When people return, with the conceptual tools and understanding gained from thinking

about other areas of philosophy, progress can be made. We figure out that going down that route no longer seems like a good idea, or we discover that certain features should be brought to bear. So, I think progress is a long-term thing. We do not necessarily answer very many questions immediately or directly, but we advance understanding in a way that allows us to do a better job when we revisit those questions.

But there will be other, new people who come back to these questions.

Maybe it is the same person. Philosophers are also very good at not digging in their heels, and they are quite good at changing their minds when the facts change. So perhaps there is also progress in individual philosophy.

Do you really think it is easy to change one's mind? There are examples like Putnam, for instance, but also many people who are really entrenched and do not move.

I hope that is not true. There are certainly some people who absolutely will not change their views, but I think the nature of philosophy is to be sensitive to evidence and argument. I think that if most philosophers are presented with good evidence and arguments suggesting they should give up a particular view, I would like to think they would do it. But maybe that is wishful thinking.

When I first started studying in London as an undergraduate, we had to do a number of exam papers at the end of the year, and I had to take one paper on Wittgenstein, Russell, and Frege. You had to answer questions on two different philosophers, and the exam rubric said that "for the purposes of this paper, the early and late Wittgenstein count as two different philosophers". So there is a nice example of someone who really did change the way they saw things.

Talking about progress and new topics, you have also written about meaning in Large Language Models (LLMs). What is your take on them?

I have written about large language models, and I do think this will be a big trend in philosophy. We used to have debates between the Turing test, which suggested we should treat a system as having rich properties like intention, meaning, and representation if it is behaviourally good enough, versus Searle's Chinese room argument, which claims that behaviour does not matter—no matter how good the behaviour of a computational system gets, something will always be missing. With large language models now, we have

systems that absolutely pass Turing tests. If you chat with ChatGPT, it can be extremely difficult to resist the impression that you are conversing with another thinking being. So, we have something capable of passing a Turing test, and now the questions are: Should we really treat them as intentional systems? Should we treat them as systems that mean or represent things? Should we treat them as agents? Do they have any kind of moral patiency? Are they entities that deserve moral consideration? All of these questions are going to become increasingly pressing as we move forward.

What is interesting about large language models is that they are black-box systems. We do not really know what they are doing or what rules they are operating under. We know that they perform next-word prediction, and somehow this gets us these fantastic results, but we do not know what the internal states of the system are really representing. I do not know whether it will be philosophers, computer scientists, or some other group that ultimately lays bare how the internal workings of these systems should be understood. Perhaps it will be a collaborative effort between all of us.

My take on them is that, for now, you should certainly think of the outputs of a large language model as meaningful because they are well-constructed outputs in a recognised, meaningful natural language. They pay attention to syntax and context and come up with appropriate, well-formed sentences. Those sentences should be treated as meaningful. However, treating the machines themselves as conversational partners is much trickier. To be a good conversational partner, I have to assume that you are mostly aiming to tell me the truth. That is not an assumption you can make about a large language model because its operating aim is not to convey truth but to convince, to say what is predictable and likely. This can diverge from what is true. Solving that problem will be very hard for programmers because truth is a very difficult notion to operationalise. My general message about large language models is: Use them, enjoy them. They do a lot of good things, but do not trust them. If the matter is important, you really need to do the checking yourself.

But when we treat them as tools, do you think it would also change the way we do philosophy or the way we read and write?

It is hard to predict what the implications of this will be. You can use these systems for crunching a lot of data, designing experiments, or ploughing through your experimental results. This could lead to projects where, in the past, you might have employed a research assistant to do that work, but now a large language model takes on the task. People will like this because it is cheaper, and funding bodies will prefer such models for cost efficiency.

However, what you lose is the unpredictability of the human element. A person's responses to tasks will be an amalgam of the things they have read, thought about, and the conversations they have had, introducing a kind of wild card element into teamwork. I am not sure you will get that same element if you replace a person with a computer.

Do you not worry that there will be a mass of philosophy books written or co-authored by ChatGPT that are much better than anything humans could write?

We should not overestimate what they do. You can push these systems over relatively easily. When they give you an answer, very often, if you then say, "Well, that is not right," they are very concessive—they will say, "You are absolutely right." I think what we want from books requires a lot more consistency over time than we currently get from these systems.

Looking ahead, what do you think will be the "hot" topics in philosophy in the near future apart from LLMs?

I do think that the kind of stuff we touched on in the lectures is likely to be interesting. There is a whole wave of experimental findings that will help us think about the skills animals and infants have or acquire, and that is likely to be quite a hot topic. I also think work in important social notions, like consent and trust, will be very important in the next few years, not least because of the impact of artificial intelligence and changes to the way we interact with others on these kinds of transactional notions. I am not sure I can think of any other topics. LLMs and the digital world are going to be such a hot topic that it will be hard for anything else to get much attention for a while.

Why should the general public be interested in the philosophy of mind? What does it offer beyond academia, and do you think it should have a broader impact?

Two things on that. First, I do think there is a place for philosophy that does not have a wider impact. If it influences other academics, that may be good enough. Sometimes philosophy has to be technical, difficult, and hard, and we should not expect the general public to understand it. On the other hand—and this is one of the things that got me into writing the book I have just written—the heuristics and biases work has had a massive public impact. If you stop a woman on the Clapham Omnibus now, she has likely heard the idea that she brings all kinds of biases to her thinking, that these are bad things, and that she is often irrational because of them. She will know the

language of the heuristics and biases programme: implicit bias and being nudged toward solutions rather than reasoning her way to them.

Part of what we ought to do as philosophers of mind is to get into that public conversation. I love that Kahneman got so many people thinking about thinking, but I believe the role of philosophy is to step in and say, "Be careful about which messages you take, and perhaps think a bit harder about thinking itself."