How Reasons Guide Us (in Reasoning and Rationalisation)

Franziska Poprawe
Independent researcher, Zürich, Switzerland
franziska.poprawe@googlemail.com

Abstract

The common-sense view that reasons guide us in thought and action and that humans are essentially reason-responsive animals is increasingly under attack by defenders of what one can call the Rationalisation View, which emphasises that we typically rationalise actions and judgements that are based on intuition rather than reasoning. This article defends the former view of human Reason, partly by replying to prominent advocates of the latter, partly by proposing accounts of reflective reasoning and rationalisation that bring to light a common, underappreciated feature: they both involve the capacity to see considerations as reasons. This capacity, the author conjectures, should be the starting point for investigating the faculty we call ‘Reason’, its evolutionary origin, and the (ir)rationality of different kinds of thought associated with it.

Keywords
reason – rationalisation – reasoning – reason module

1 Introduction

It is an integral part of folk psychology and common wisdom that when we make up our minds about what to do or what to believe, we (should) attend to the reasons we have for doing or believing one thing rather than another. In moral philosophy and epistemology alike, it has long been recognised as a hallmark of reasons that they guide us in thought and action (Grice 2001; Scanlon 1998; Dancy 2002; Korsgaard 2009; Hieronymi 2013; Lord 2018; Brandom 1982; Anscombe 2005; Alvarez 2009; Rinard 2019). We reason correctly in so far as
we follow the path of reasons correctly (Piller 2001; Dancy 2018). Indeed, the
capacity to understand and respond to reasons as such is seen as a unique
feature of human nature (Hacker 2007; Korsgaard 2008): ‘In a world without
reasons, we would act only on our instincts and desires, living as other an-
mals live’ (Parfit 2011b, 620). Accordingly, it is thought since Aristotle, reasons
play a central role in understanding the arguably distinctively human faculty
of ‘Reason’.

This paradigm is increasingly under attack from psychologists and philoso-
phers who argue that reasons are not there to guide our conscious inferences
but to rationalise our intuitive judgements and decisions. While they don’t
deny that humans can respond to reasons, they argue that reasons do not play
the dominant role in human cognition that they are traditionally assumed to
play, and that, accordingly, models of human Reason should not focus on con-
scious reasoning on the basis of reasons but instead on intuitive thought pro-
cesses, that is, associative thinking that is at least partly unconscious (Mercier
& Sperber 2017, 7). I call this general view, which various theorists defend in
different versions and contexts, the Rationalisation View.

Jonathan Haidt (2001), for instance, famously defended a model of moral
reasoning according to which ‘moral reasoning is an effortful process, engaged
in after a moral judgment is made, in which a person searches for arguments
that will support an already-made judgment’. So moral judgements, on this
social intuitionist model, are driven by partly unconscious, intuitive, emo-
tional processes which are not guided by the conscious attention to reasons,
but which are then post-hoc rationalised. Recently, Cushman (2020) proposed
a theory of rationalisation that emphasises its ubiquity and adaptive fitness.
Rationalisation, he argues, ‘takes an action that has already been performed
and then concocts the beliefs or desires that would have made it rational’. It is
not only ‘designed to infer the underlying causes of our behaviour for the sake
of explanation’ but ‘to extract information from the non-rational processes
that influence our behaviour (ibid, 2). Most extensively, Mercier and Sperber
(2017) provide conceptual and empirical arguments for the thesis that the
‘main role of reasons is not to motivate or guide us in reaching conclusions but
to explain and justify after the fact the conclusions we have reached’ (Mercier
& Sperber 2017, 112). Their view is that reasons do not primarily guide us in
decision-making and judgement but that reasons are for ‘social consumption’
tools for managing one’s reputation, making public commitments, and attain-
ning a useful social standing. Accordingly, on this view, conscious ‘reasoning
proper’ or ‘reflective reasoning’ is not to be understood as a distinctive kind
of slow thought guided by our recognition of reasons but as just another kind
of intuitive inference. In contrast to Haidt, Cushman, and Kahneman (2011),
amongst others, Mercier and Sperber do not only emphasise the importance of intuitive thinking (system-1) as opposed to reason-guided deliberation (system-2) but even deny that there is anything but intuitive inference, that there are two fundamentally different types of thinking.

The first aim of this article is to defend the common-sense view that we are rational animals by virtue of recognising and responding to reasons and that, accordingly, reasons guide us in thought and action. These features are essential and central to human Reason. I defend this view against influential counterarguments recently put forward by the mentioned proponents of the Rationalisation View. I draw attention to three inter-dependent questions, which – directly or indirectly – drive the current, interdisciplinary debate on Reason:

1. What forms of reasoning or inference can be distinguished?
2. What is the role of rationalisation in human thought?
3. What are reasons for – individual guidance, social exchange, or rationalisation?

I argue that the answers that some defenders of the Rationalisation View give to these questions are implausible or unsupported. Furthermore, I point out that, while it is generally uncontroversial that our answers to these questions contribute to our understanding of Reason, it is rarely discussed how exactly they do so and why they answer genuinely different questions. What do they mean and how do the concepts of ‘reasons’, ‘reasoning’ and ‘rationalisation’ relate to ‘Reason’? Therefore, the second aim of this article is to clarify questions (1)–(3) and, based on that, propose a novel approach to Reason.

I proceed as follows. Section 2 gives a summary of Mercier and Sperber’s approach to Reason and reasoning. Because I consider their theory the most elaborate and strong defence of the broader position I wish to address, I formulate arguments for the distinction between reasoning and intuitive inference primarily in response to their specific proposal (question (1)). Section 3 is more constructive. It discusses the nature of rationalisation and its relation to Reason, giving a novel answer to question (2). The analysis leads me to the following conjecture: The capacities to see things as normative reasons for certain responses and to ‘distinguish what is a reason from what is not, and to respond accordingly’ (Dancy 2018, 143) are essential to Reason. I show how these capacities are manifest not only in reflective reasoning but also in various types of

---

1 Mercier and Sperber (2017, 7, 144) seem to treat them almost interchangeably. The study of human Reason is assumed to be the same thing as the study of reasons, as well as the study of reasoning.
rationalisation. The analysis allows us to see what is and is not rational about rationalisation. In Section 4, I turn to question (3) which, I argue, is misleading and in need of reformulation. The proposed approach to Reason turns out to sit well with the view that the function of Reason is primarily social and the view that Reason is a group-level adaption. Section 5 concludes.

2 Is Reason a Module?

It is often assumed that the capacity called ‘Reason’ is inherently related – if not identical – to the capacity to reason, that is, the capacity to engage in structured, conscious deliberation about meaningful contents. Following this trend, Mercier and Sperber (henceforth M&S) investigate Reason by first investigating reasoning. Their answer to question (1), in short, is this: There is one form of inference, namely intuitive inference. What we call reasoning is intuitive inference about reasons. This kind of intuitive inference is what defines Reason, which is thus best understood in terms of the ‘reason module’, as a cognitive inference mechanism that produces intuitive inferences about reasons.

After presenting their account of intuitive inference and reasoning in more detail, I raise several objections to it and show how it contradicts the common-sense view. I argue that we cannot reduce the mental activity of reflective reasoning, the kind that involves attention to reasons, to the working of an inference mechanism (2.1). Further, I doubt that we have grounds to believe that there is a ‘reason module’ (2.2). I end this section by questioning the claim that intuitive thinking is central to Reason (2.3).

2.1 Reasoning

M&S define inferences in general as processes that extract ‘new information from already available information’ (M&S 2017, 52–53). There are different kinds of inferences or different ways of performing inferences. On the one hand, there are intuitive inferences, automatic processes of which we are not conscious. The outcomes of intuitive inferences are intuitions. We experience them as ‘something our mind produced but without having any experience of the process of its production’ (ibid, 65). On the other hand, there is ‘reasoning proper’, or ‘reflective reasoning’ (from now on: reasoning), which is a special way of performing inferences that is conscious and involves attention to reasons. When we reason, ‘conclusions do not just pop up in our mind as self-evident; we arrive at them by considering reasons to accept them’ (ibid, 52). While that much seems acceptable to many, M&S aim to give a new theory of reasoning in order to replace the dual-process theory, which draws a
sharp line between two kinds of thinking, system-1 and system-2 (Stanovich and Evans 2013; Kahneman 2011). Reasoning is often thought of as a paradigmatic case of system-2-thinking and contrasted with the intuitive, automatic processes of system-1-thinking. M&S deny that there are two fundamentally different ways or systems of thinking and hold that reasoning is – in the end – just a kind of intuitive inference. Let me illustrate their account with the help of the following example. Suppose you reason:

\[
P \quad \text{‘The pavement is wet.}
\]
\[
C \quad \text{So, it has been raining’.}
\]

According to M&S, the conclusion that it has been raining is an output of the so-called ‘reason module’, which is ‘a mechanism of intuitive inference about reasons’ (M&S 2017, 107). Another output of this mechanism is the intuition that the fact that the pavement is wet provides a reason to believe that it has been raining. Thus, reasoning involves a higher-order intuition about how lower-order intuitions (here: that the pavement is wet) support a conclusion (ibid, 166). In general, reasoning is nothing but this twofold output of the reason module. In general, an instance of reasoning consists of (ibid, 149)

(i) an intuitive, spontaneous inference based on a set of premises \(P\), yielding the lower-order conclusion, \(C\), which is a new piece of information about the world, and

(ii) a higher-order intuitive inference about reasons, yielding the higher-order conclusion that \(P\) provides reasons for accepting \(C\).

There is, therefore, no room for system-2 (ibid, 153). Like any cognitive module, it has a history, function, and procedure, taking certain representations of facts as inputs to draw conclusions from them (ibid, 118). The specific function of the reason module is ‘the identification of reasons’ (ibid, 144).

This brief but I hope accurate summary of M&S’s elaborate theory must suffice for the present purpose. Again, I believe some of its fundamental assumptions and claims are shared by many, in particular:

(a) Some of our beliefs (and actions) about the world are the outcomes of processes that we are not conscious of, like the processes underlying perception, association, or remembering. M&S call these attitudes ‘intuitions’, and the underlying processes ‘intuitive inferences’.

(b) There is such a thing as reflective reasoning. It is conscious throughout: we are not only aware of the conclusions we draw but also of the reasons
for them. Reasoning essentially involves *recognising* reasons and appears to be uniquely human (ibid, 5, 144).

(c) We make judgements about reasons. We hold that some considerations are reasons to believe or do certain things and that some considerations are stronger reasons than others. Some of those judgements are arrived at through reasoning, others through intuitive processes.

Despite this common ground, I would now like to raise two major doubts about M&S’s view of reasoning. First, I believe it is explanatorily insufficient.

Reflective reasoning, M&S argue, is nothing but the twofold output of the reason module, yielding outputs (i) and (ii). Later they add that C is “a **reflective conclusion**, a conclusion accepted because of higher-order thinking (or “reflection”) about it” (ibid, 150, italics in the original). The intuition of (ii) “provides a reason for what is, therefore, a reflective conclusion” (ibid, 150). With this, the account says the following: the process that produces the conclusion C counts as an instance of reflective reasoning if and only if the agent comes to accept C *because of* (ii).

While I believe this relation between (i) and (ii) is plausible and corresponds to a widely accepted claim in the contemporary philosophical literature of reasoning, a full account of reflective reasoning would have to further explain this relation between (i) and (ii), as is also widely recognised in the debate. To the first point, the thought is that in reasoning agents come to adopt conclusions *because* of their recognition of the normative reasons for them, that is, because they see some considerations as speaking in favour of adopting them. Reasoning is not simply coming to adopt some conclusion while also recognising that some facts are normative reasons for accepting that conclusion. We make reasoned judgments and decisions **in light of** and **for** reasons. Reasoning is the paradigm case of using reasons prospectively, of responding to (normative) reasons **for** a certain kind of response. This widely accepted point has generated a heated debate on the question of what exactly the (causal) role of the reasoner’s **recognition** of reasons in the reasoning process is (Boghossian 2014; Raz 2015; McHugh and Way 2016; Kauppinen 2018). In contrast, M&S offer no answer to this question. They do not explain the role of (ii) in the process leading to C. Further, it appears to me that the kind of explanation that their account would allow is implausible: Since both, (i) and (ii) are said to be outcomes of the very same reason module, M&S would have to say that one outcome of the module explains the other, and this seems odd.

The second problem I see with the account concerns (ii). The higher-order outputs of the reason module are said to be intuitions ‘about reasons and the support these reasons give’ (ibid, 149). In the example above, the higher-order
output of the reason module would be the belief that the fact that the pavement is wet is a reason for (or supports) believing that it has been raining. With (ii), M&S account for the widely accepted idea that reasoning involves recognising reasons. They thereby employ a particular understanding of the latter: ‘Only if you were to entertain a thought like ‘From the fact that the pavement is wet it follows that it must have been raining’ would you be recognising the reason for your conclusion’ (M&S 2017, 131, my emphasis). This substantive account of what it is to recognise a fact as a normative reason is, I believe, implausibly strong. It implies that explicit beliefs about what is a reason for what are necessary for reasoning, which seems false. Rational human adults (and children from a certain age) can engage in reasoning about what to believe or what to do without thereby employing sophisticated concepts like ‘reason for’, ‘support’, or ‘implication’. I can derive the claim that it has been raining without forming an explicit belief about what is a good reason for believing it. In fact, this point is also widely accepted in the philosophical literature (see references above). Reasoning, it is agreed, involves recognising reasons, taking things to be reasons, or seeing one’s premises as reasons. It is, ex hypothesi, drawing conclusions in response to the normative pressures we are sensitive to. However, reasoning does not necessarily involve explicit beliefs about reasons, for this would render reasoning an implausibly sophisticated exercise.\(^2\) What exactly it is to recognise reasons or see one’s premises as reasons is debated but doxastic accounts of these notions, according to which they refer to beliefs about reasons, have been rejected. M&S correctly explain that recognising some fact \(f\) as a reason to do or believe something (to \(\varphi\)) does not only consist in believing that \(f\) is a reason to \(\varphi\).

Moreover, there is another way in which M&S’s account of recognising reasons is implausibly strong. The claim that ‘only if you were to entertain a thought like ‘From the fact that the pavement is wet it follows that it must have been raining’ would you be recognising the reason for your conclusion’ (M&S 2017, 131, my emphasis), it seems to follow that we can only recognise pro toto reasons, that is, reasons that warrant drawing a certain conclusion. But there are also pro tanto reasons, reasons which count in favour of a certain response but may be outweighed and thus not warrant that response, all things

\(^2\) Such a belief rarely plays a role in individual reasoning as well as in our evaluations of the arguments given by interlocutors. Children can engage in arguments, give reasons and evaluate others’ reasons without applying the concept of ‘a reason’ (see for example Köymen and Tomasello 2020; Köymen, Mammen, and Tomasello 2016; Castelain, Bernard, and Mercier 2018). This suggests that they can do the former before the latter.
considered (Broome 2004). When we recognise pro tanto reasons for our conclusions, we certainly do not (need to) believe what M&S claim we must. For example, I may recognise the fact that the neighbour said it would be raining soon as a reason to believe it will be raining soon, without believing that from the neighbour’s claim (alone) it follows that it will be raining.\(^3\)

In sum, I have argued so far that M&S do not sufficiently explain the role that recognising or seeing things as reasons plays in reasoning, nor what the former consists in. The common-sense view, according to which reasoning involves recognising the premises we adduce as reasons for our conclusions, remains undefeated. Importantly, it is not necessarily committed to the implausible claim that reasoning necessarily involves beliefs about reasons, like the belief that the wetness of the pavement provides a reason to believe that it has been raining. My analysis suggests that because M&S adopt (without argument) an implausibly strong account of what it is to recognise reasons, they arrive at an implausibly strong account of reasoning that renders reasoning a sophisticated, conceptually demanding, hence rare exercise. This, of course, supports their focus on intuition and rationalisation. However, I believe there is no good reason to accept their strong, doxastic account of recognising reasons. It is not straightforward to spell out what recognising or seeing things as normative reasons is (I now stick to the latter notion). From this insight, we must not conclude that reasons do not usually guide us but rather that it is hard to specify how they guide us in reasoning.

2.2 Is There a ‘Reason Module’?

M&S’s (ibid, 144) proposal leads them to the speculation that ‘the closest thing to classical “Reason” to be found in the human mind/brain may well be’ the reason module. Even if this module cannot precisely correspond to that which ‘Reason’ traditionally refers to because the latter refers to a feature of the whole mind rather than only one mechanism, it is supposed to get us a long way in understanding Reason. So, on what grounds can we believe that we are equipped with an automatic mechanism with the function of recognising reasons?

For the sake of argument, let us accept M&S’s understanding of ‘recognising reasons’ as coming to believe, through inference, that something is a reason. Because this inference must at least sometimes be intuitive and not based on reasons, they argue, it must at least sometimes be the product of a module, an automatic mechanism. Thus, there must be a module with the function of

---

\(^3\) I thank an anonymous reviewer for pointing this out to me.
inferring reasons in this way. This is the reason module. The argument that M&S give for the first central claim of this argument, namely that the recognition of a reason must at least sometimes be due to an intuitive inference, is the following (ibid, 131–132):

So, how are reasons inferred? By finding further reasons for our reasons? Sometimes, yes; most of the time, no. Assuming that the recognition of a reason must itself be based on a higher-order reason would lead to an infinite regress; to infer a conclusion A, you would need some reason B; to infer that B is a reason to infer A, you would need a reason C; to infer that C is a reason to infer that B is a reason to infer A, you would need a reason D; and so on, without end. Hence, the recognition of a reason must ultimately be grounded not in further reasoning but in intuitive inference.

M&S attribute this argument to Lewis Carroll (1895), even though his original argument is not about how we derive reason-claims, but about the role of formal inference rules in deductive reasoning and the question of whether accepting the conclusion because of the premises is something over and above accepting the premises and accepting the conclusion. In any case, I hope the following is a charitable reconstruction of the cited argument: To say that we always infer reason-claims, like the claim that B is a reason for believing A, through reasoning would lead to an infinite regress. Hence, at least some of our reason-claims are the outcomes of intuitive inference. Hence, there must be a reason module that has the function of inferring reason-claims, that is, the function of ‘inferring reasons’.

I believe the first premise is false. ‘Inferring reasons’ through reasoning, in the sense in which M&S use this expression here, does not lead to an infinite regress. To see why, let us think through the quoted argument with exemplary statements instead of the placeholders A, B, and C. Take as an example the inference from B to A:

(B) The pavement is wet.
(A) So, it has been raining last night.

4 Note that with respect to the job of the reason module, the expression of ‘inferring reasons’ is used with two different meanings. Sometimes ‘inferring reasons’ stands for inferring that some fact is a reason (that is, inferring what I call a ‘reason-claim’); sometimes, in the context of rationalisations, it stands for collecting and naming considerations as alleged motivating reasons for one’s beliefs or actions in order to justify them (M&S 2017, 140).
How could we ‘infer that B is a reason to infer A’ (ibid, 131–132)? We could say:

(C) That the pavement is wet is evidence (makes it likely) that it has been raining.
(R) Therefore, that the pavement is wet is a reason to believe that it has been raining.

Now we might ask: why is C a reason to believe R? And we might answer: because we have reason to believe what the evidence suggests. In the next iteration, we might then ask: why do we have reason to believe what the evidence suggests? And this is a substantive normative question with no easy answer. A possible answer is: because we ought to believe the truth.

One might disagree that C is the correct answer to the question of why the pavement being wet is a reason to infer that it has been raining. So, consider an alternative approach to justifying the inference from B to A. Suppose it proceeds on the background assumption (B*) that the pavement is wet only if it rained. We could then say:

(C*) That the pavement is wet, together with the fact that the pavement is wet only if it rained, logically implies that it has been raining.
(R) Therefore, that the pavement is wet (together with the fact that it is only wet when it rained) provides a reason to believe that it has been raining.

Now we might ask: why is C* a reason to believe R? And we might answer: because we have reason to believe what is logically entailed by known facts. In the next iteration, we might then ask: why do we have reason to believe what is logically entailed by known facts? Again, we have reached a substantive normative question that calls for theorising about the norms of belief, and why it is right to believe the truth.

So, no matter which line of thought you accept, the structure of giving the reasons for why some facts are reasons (in M&S’s terms: the structure of deriving reasons from further reasons) is the same. It does not lead us to an endless regress, but – to the contrary – to rock bottom. It leads to fundamental normative claims. Our explanations of why some considerations are reasons must come to an end, as any non-circular explanation must.

I agree that the fundamental normative claims we eventually reach may be blunt intuitions that we cannot explain and cannot derive through reasoning from other claims we accept. The reason-claims we derive through reasoning are then, down the line, based on intuitions that are not derived through
reasoning. However, these basic intuitions are not intuitions about what is a reason to do or believe something but intuitions about the much broader questions of what is valuable, desirable, right or good. The mechanism that produces them is not a mechanism with the function of deriving reason-claims like R. It is not that which M&S call the reason module. Instead, it is a mechanism that produces intuitions about what is good or bad, and thus a sense of how things are to be done. What underlies our capacity to derive reasons through reasoning is the general capacity for normative thought and values (more on this in Section 4).

Of course, in addition to reflectively deriving reason-claims from what we take to be true, we can, and often do, judge intuitively without reasoning that some fact is a reason to do or believe something (this is point c) above). But this finding lends no more support to the postulation of a reason module than the finding that we sometimes make intuitive judgements about the temperature outside lends support to a weather module. Further, there are indeed grounds to believe that intuitive thought processes must play some role in reflective reasoning. But this does not imply that reasoning is nothing but intuitive inference. That something is a constitutive part of something else does not reduce the latter to the former.

So, I see no grounds to attribute to humans a reason module, nor to understand the peculiarities of thought about and with reasons in terms of a module. In addition to the remarks I made, I see the following reason not to do so. Upholding the distinction between, on the one hand, reflective reasoning and, on the other hand, automatic intuitive processes run by modules allows us to accommodate an important assumption, namely that reasoning is something we do, something we intentionally and consciously engage in, and not something that merely happens to us. We cannot explain what reasoning, understood as a person-level activity, consists in if we investigate how a sub-personal ‘module recognises reasons’ (M&S, 132), why it ‘behaves the way it does’ (M&S, 9). We must investigate what we have to do with it.

5 Dutilh Novaes (2018, 517) raises further worries, compatible with mine, about the existence of a specialized reason module.

6 While this conceptual distinction supports dual-process theories, it is not clear how it relates to the imprecise distinction between system-1 and system-2 thinking (Stanovich and Evans 2013; Evans 2019). I see two main hypotheses. One could hold that active, person-level reason-responsive reasoning is either (i) the prototypical system-2 thought process and qualitatively different from system-1 processes, or (ii) a special kind of system-2 process that is qualitatively different from both system-1 processes and other system-2 processes.
2.3 What Kind of Thinking Manifests Reason?

Defenders of the Rationalisation View tend to emphasise the ubiquity of intuitive inference and rationalisation (for example, Cushman 2020). M&S (2017, 143) even conclude that we are not guided by reasons, that ‘the presumption that people’s beliefs and actions are motivated by reasons is empirically wrong’. They argue (ibid, 142):

Contrary to the common-sense view, what happens is not that we derive intuitive conclusions from reasons that we somehow possess. What we do, rather, is derive reasons for our intuitions from the intuitions themselves by a further process of backward inference.

When we rely on our intuitive inference mechanism to guide us through the day, then, by definition, we do not reason to our conclusions by deriving them from reasons. But this does not imply that we never derive conclusions from reasons, nor that we do not have the capacity to do so. The fact that we are necessarily unaware of the mere causes of some attitudes does not rule out that there are also other attitudes that are not the result of merely causal mental processes or automatic mechanisms. The limits of backwards introspection do not constrain mental action forward. Some attitudes are the conclusions of active reasoning on the basis of premises that we see as reasons for us to respond to the world in one way rather than another. Often enough our decisions are the results of reasoning processes (Stafford 2015; Dahl et al. 2018; Dahl and Waltzer 2020).

---

7 In this article, I distinguish between two types of thought: intuitive inference mechanisms and reflective reasoning. More precisely, I am concerned with intuitive inferences that are not reason-responsive, and reflective reasoning that essentially involves responding to reasons. However, I do not claim, vice versa, that responding to reasons necessarily requires reflective reasoning. We sometimes count as responding to reasons even when we do not actively derive our conclusions through reasoning. That is, some thought processes do not qualify as active, reflective reasoning, but nevertheless manifest our capacity to respond to reasons. I argue elsewhere that these processes are different from the non-reason-responsive, sub-personal, intuitive thought processes which I am concerned with here. The main difference is that the former are under the agent’s virtual control (Pettit, 2007a, 226-267): they are accessible and (with effort and reflection) re-constructible. For example, when you habitually walk home from work without reasoning about the turns you take, you respond to the reasons you have to take these turns without reasoning about them. Or take the example of Huckleberry Finn: he may count as unreflectively responding to the reasons to help Jim, even though he believes that he is not responding to reasons (Arpaly 2003, 75-77; Mantel 2018, 83). The possibility of responding to reasons without reasoning – through habit, skill or virtue – complements the distinction drawn in this article, namely that between sub-conscious, intuitive inference mechanisms on the one hand, and reflective reasoning on the other.
The more one observes how people intuitively jump to conclusions and rationalise those, the more tempting it may be to conclude that this behaviour is a *typical* manifestation of Reason. But that we often rely on intuitive or habitual inferences does not imply that we ‘typically construct our reasons after having reached the conclusions they support’ (ibid, 142, emphasis added). Importantly, it does not imply that this is the *typical* – or *central* – feature of Reason. On the contrary, one could argue that some cases in which we fail to attend to and consciously consider the reasons we take ourselves to have, cases in which automatic inferential processes are left to work by themselves, are *not* instances of the manifestation of Reason at all. They are cases where Reason should be employed but is not – at least not appropriately. These cases should then not be our focus when studying the faculty of Reason. Of course, understanding these cases is interesting in and of itself. Why is Reason ‘presumably’ not employed in the experiments M&S discuss (ibid, 121)? That is, why do people seem to react without reflection and active deliberation in cases in which we think they should?

The point is that one must carefully distinguish between at least three questions. First, a conceptual question is what it is for a certain type of thinking to take place. A second, also conceptual question is whether a certain type of thinking and the subsequent behaviour manifests the faculty we call Reason, and in what way. And a third, empirical question is when and why which type of thinking occurs. Once we have answered these questions, we can judge whether a certain type of thinking is a manifestation of Reason and whether it is an adaption or a bug (see also Section 4).

To summarise, I have here defended the answer that the common-sense view gives to question (1) against Mercier and Sperber’s proposal to understand Reason in terms of intuitive thought processes. Specifically, I objected to their account of what it is to *recognise* or *see things as reasons*, as well as their argument for a reason module. I emphasised the central role of the capacity to *see things as reasons* in reasoning. In the next section, I argue that this capacity is also essential to rationalisation. The analysis will lead to the (arguably Aristotelian) conjecture that this capacity is at the core of Reason: we take ourselves to have normative reasons for responding to the world in certain ways.

---

8 Though Kahneman is often said to have contributed substantially to our view of ourselves as irrational animals driven by biased intuition, I believe what I say here is compatible with much of his work. For studying the ways in which intuitive processes are ubiquitous and flawed has not led him to conclude that they are essential to Reason but – *au contraire* – to the hypothesis that “the voice of reason may be fainter than the loud and clear voice of an erroneous intuition” (Kahneman 2011, 417). I am here concerned with our approach to the former, not the peculiarities of the latter.
3 Is There Reason in Rationalisation?

As the name implies, the Rationalisation View holds that rationalisation is a well-adapted, key feature of Reason, while common sense tells us that rationalisation is a kind of failure of Reason. In this section, I address this controversy and provide a new answer to question (2): What is the role of rationalisation in human thought? To do so, I look at different kinds of rationalisation, clarifying what kind of capacities they require, how they relate to Reason, and what is and is not rational about rationalisation.

3.1 Types of Rationalisation

According to a widespread conception, rationalisation is the practice of adopting or at least reporting the beliefs and desires that would have made an already performed action or judgment rational. This conception underlies several theories of rationalisation, which give different explanations for why people engage in this practice. It is, for instance, the basis of Festinger’s (1957) classical theory of cognitive dissonance, according to which people are motivated to post-hoc rationalise their actions and beliefs by a desire to avoid the psychologically aversive state of cognitive dissonance. Also, as we have seen, M&S employ this conception of rationalisation, emphasising that people rationalise their judgements and actions in order to convince others or to appear in a positive light (see also Haidt 2001). Similarly, to give a last recent example, this conception of post-hoc rationalisation is at the core of Cushman’s (2020, 3) theory of rationalisation, which holds that the function of rationalisation is to extract information from the intuitive processes that influence our behaviour in order to use it in future reasoning.

Leaving the differences between these theories of rationalisation and the question of its function aside, we must note that the term ‘rationalisation’ may also refer to a quite different phenomenon. On the assumption that it is plausible to speak of multiple kinds of rationalisation rather than competing accounts of what rationalisation is, let us call this phenomenon self-serving rationalisation (Quilty-Dunn 2020). A paradigmatic example is the following case:

Tom wonders whether to leave the office early today to have dinner with his friends, who spontaneously invited him. Until he received the call, he thought he ought to keep his promise to his team and finish his task for the joint project by the end of the day. But now he reasons as follows: ‘I haven’t seen my friends for a while and who knows when they all have time again. The project won’t be delayed if I finish the task tomorrow, so the team shouldn’t mind. Anyway, breaking one promise is not so bad. So, I shall go to the dinner’. 
Tom rationalises his decision. But he does not post-hoc rationalise it. Instead, he engages in a kind of motivated reasoning that is distorted by self-interest. An underlying motive – the desire to go to the dinner – distorts the direction of his reasoning while it remains hidden throughout the thought process (Schwitzgebel and Ellis 2017). The motive influences the deliberation, driving the decision to go to the dinner even though it does not provide a reason for going.

So, there are at least two types of rationalization, post-hoc and self-serving rationalisation. What do they have in common? Are they adaptive, as Cushman argues? When contrasting the two, Cushman (2017, 44) speculates that they may ‘share a common mechanism’. I propose to first better understand the nature of each type of rationalisation in order to then see a) what unifies them, b) in what sense they are or are not rational, c) how they relate to Reason, and d) whether they have an adaptive function.

3.2 Seeing Things as Reasons, Again
Consider the following case of post-hoc rationalisation:

Sara is at the store and absent-mindedly puts two packages of chocolate bars on the conveyor belt while queuing. Back at home, her partner asks her: ‘Why did you buy so many chocolate bars?’ She replies: ‘Because we are going hiking next weekend, remember? And they are my dad’s favourite’. Her partner nods convincedly, knowing that the chocolate bars have proven to be of high value on their hiking trips and that Sara’s dad is coming to visit next week.

Let us suppose that Sara really did not consciously reason about whether to buy the chocolate bars and mindlessly put them on the conveyor belt, say, driven by hunger and a desire for sweets. How can we describe her rationalisation in response to her partner? According to Mercier and Sperber (2017, 142), when we post-hoc rationalise our actions, we ‘derive reasons for our intuitions from the intuitions themselves by a further process of backward inference’. According to this view, Sara derives the reasons that rationalise her intuitive action from the intuition itself through backward inference. But how could Sara infer the reasons she gives to her partner simply from the consideration that she did in fact buy the chocolate bars? It seems false, or even nonsensical, to say that Sara derives the proposition that ‘we are going hiking next weekend’ from the recognition that she bought chocolate bars. To come up with the reason, she must rely on certain cognitive skills, most notably memory and association. She must remember the hiking trip, that the chocolate bars have been proven valuable on hiking trips, that her father is coming to visit, etc. But
remembering these facts is still not enough to count as having derived reasons. This is so because thinking that something is the case is not the same as entertaining that consideration as a reason (M&S 2017). Again, the latter does not require the explicit belief that something is a reason but it must involve seeing the consideration as a reason (see Section 2). When Sara explains her action, she sees the considerations she mentions not only as explaining her behaviour but as justifying it. She sees them as normative reasons for buying chocolate bars. Post-hoc rationalisation, just as forward-looking reasoning, essentially involves the capacity to see ordinary facts about the world (like the fact that someone is visiting) as normative reasons for us to do things, as favouring or even requiring a certain kind of response. Therefore, reasons play a guiding role in rationalisation: how we rationalise our behaviour depends on the reasons we see for it. Rationalisation is not to be understood as a kind of backwards inference or extraction, but as the process of picking out those features of the situation which we see as reasons for our actions or beliefs and communicating them accordingly.

In the case of Sara, where rationalisation is embedded in a social setting, there is another aspect to mention. Sara successfully rationalises her action to her partner because he also sees the considerations she gives as reasons for buying the chocolate bars. When responding, Sara must factor in what kind of reasons her partner is likely to accept. In general, it is not just that ‘we would like others to be guided by the reasons we give them’ (M&S 2017, 113, my emphasis). We would treat our interlocutors differently if we did not assume that they will be equally guided by the reasons we give them.

So, post-hoc rationalisation is another manifestation of the capacity to see things as reasons, whereby both, those who justify themselves by giving reasons and those at the receiving end must manifest this capacity for the practice of rationalisation to fulfil its function. In post-hoc rationalisation, people are guided by (what they see as good) reasons for certain attitudes and communicate those reasons accordingly. Those who successfully rationalise their actions in social settings are therefore rational in the following sense: they correctly employ their sensitivity to reasons, even if only after the fact. They treat features of their situation as reasons, which really are reasons (Velleman 1996). They manage to point to those features of their situation which really do provide reasons for their actions in the given context. If these reasons are also good moral reasons rather than, say, good prudential reasons, then post-hoc rationalisation also has a moral value: it forces us to reflect on what counts in favour of what and commits us to openly endorse the underlying moral principles.9

9 I focus on post-hoc rationalisation for the purpose of social justification rather than (solitary) self-explanation here but the latter is arguably also practically and morally beneficial due
Why does rationalisation have such a bad record, then? Paradigmatically, people post-hoc rationalise those behaviours or judgments that they arrived at intuitively, without engaging in reasoning. Since the intuitive processes underlying those ‘automatic’ responses to the world are opaque, when we explain or justify the intuitions they lead to, we can only make up stories about their causal histories. The beliefs and desires we adopt or at least report in post-hoc rationalisation are not the motivating reasons underlying our actions or beliefs but confabulations (Nisbett and Wilson 1977). Instead of admitting ignorance of the causes of our behaviour, we pretend that we did reason about what to do or believe. This pretending may not be deceptive, however. Those who confabulate rationalisations may be unaware of their ignorance, believing that they truthfully report how they made their choices or judgments with no intention to deceive (Coltheart and Turner 2009). But even if ignorance offers some excuse, post-hoc rationalisation may still be epistemically objectionable, namely when the rationalising agent has no good evidence for the alleged basis of their attitudes (Summers 2017; Bortolotti 2018), and yet adopts the belief that the action or judgment in question is based on the beliefs or desires that make it rational. In any case, what seems much more objectionable are cases of post-hoc rationalisation that involve an outright, intentional lie. These are cases where an agent does not rationalise an intuitive behaviour that they need to explain to themselves or to others but a well-thought-through, self-serving action, the reasoning behind which they wish to conceal. Suppose Sara did in fact contemplate whether to buy the chocolate bars and bought them because she wanted to treat herself to a movie and sweets tonight. Ashamed that she broke her resolution to reduce late-night snacking, she pretends to have bought the chocolates with a nobler purpose in mind.10

So much about post-hoc rationalisation. Turning now to self-serving rationalisation, consider the case of Tom again. First, note that Tom also manifests the capacity to see things as reasons. He carefully weighs considerations that speak for and against accepting the dinner invitation and makes a deliberate decision based on that. The issue is that Tom does not employ the capacity to its effects on perceived agency (Bortolotti 2018), self-consistency (Velleman 2000), and first-person authority (Summers 2017).

---

10 This suggests that people may be differently motivated to rationalise their behaviours. If Sara is motivated to rationalise her behaviour by the desire to appear in a better light (together with the belief that rationalising allows her to appear in a better light), this is her motivating reason for the intention to rationalise, which in turn leads her to search for the best reasons to do so. Her motivating reasons to rationalise do not directly guide her search for reasons in the rationalisation, but rather the preceding decision-making as to whether to rationalise her action in the first place.
to respond to reasons correctly. He ignores, or at least discredits, reasons that speak in favour of working tonight, and gives more weight than he should to the reasons that speak in favour of meeting his friends. His desire to go to the dinner distorts his reasoning and decision making. Because he reasons, he is subject to the respective evaluative standards, according to which he reasons incorrectly in the sense that he responds incorrectly to the normative features of the situation. He treats the considerations he adduces as providing conclusive reason to go to the dinner while in fact they don’t. He is thus irrational in the sense used above. Further, Tom’s behaviour appears morally objectionable. His self-interest drives him to accept dubious moral principles. For example, he downplays the importance of promise-keeping in way a that he would probably otherwise not.11

In conclusion, both types of rationalisations share a common feature. They require the capacity to see things as reasons.12 This, together with the discussion of reflective reasoning in Section 2, leads to the conjecture that this capacity to see things as normative reasons is essential to and a key feature of Reason. It is the capacity that allows us to not only see the world as it is but also to appreciate what it demands from us, and to be sensitive to normative pressures. The conjecture implies that Reason is involved in post-hoc rationalisation as well as in self-serving rationalisation. Where the former is rational in the sense that it is a correct exercise of the capacity to see things as reasons (as long as it does not involve an intentional deception or motivated reasoning), the second is irrational because Reason is engaged but in an inappropriate way. Self-serving rationalisation is a type of incorrect reasoning: we wrongly treat considerations as reasons for certain actions or give them more weight than we should. Importantly, to say that Reason is manifest in rationalisation is not to say that rationalisation is a defining feature of Reason, as the Rationalisation View holds. Again, the proposal is that the capacity to see things as reasons is essential to Reason, which is involved in reasoning as

---

11 Post-hoc rationalisation and self-serving rationalisation may fall together, bringing the agent to lower their moral standards for the sake of attitudinal consonance. This happens when an agent attempts to post-hoc rationalise an action that is unjustified on higher moral standards, and can only be justified by tweaking moral principles, for example, by changing one’s belief that promises ought to be kept to the belief that it is ok to break one promise once in a while (see also Jefferson (2020)).

12 Several arguments in the debate on the practical and epistemic benefits of rationalisation allude to reason-responsiveness but do not explain how exactly the capacity to see or take things to be reasons plays a role in rationalisation (Velleman 2000; Moran 2001; Strijbos and de Bruin 2015; Summers 2017; Bortolotti 2018).
well as in rationalisation. Some instances of rationalisation are rational in the sense explained above, others are not. I take this result to be an elaboration of the common-sense view.

The proposed conceptual approach to Reason does not yet provide an answer to the question of why we tend to employ Reason in some situations but not others. Why do we post-hoc rationalise our actions in ways that portray us in a favourable light? Why do we sometimes fail to manifest Reason appropriately and engage in incorrect, self-serving reasoning? In short, why did Reason evolve the way it did? This brings me to my final topic of investigation, namely the third of the three questions raised at the outset.

4 What Are Reasons for? What Is Reason for?

In the old paradigm, reasons are there to guide us in thought and action, while the Rationalisation View, in contrast, holds that reasons serve to post-hoc rationalise our intuitions. So: ‘What are reasons for – individual guidance, social exchange, or rationalisation? The conjecture I reached above, that the capacity to see things as reasons is central to Reason, points to the following answer.

Reasons play a prospective role in reflective reasoning, where the conclusions we draw are based on the reasons we take ourselves to have, as well as a role in retrospectively justifying our actions and judgements. Can we say that one function takes priority? I believe not. Even if most of our actions and judgements in daily life were the result of intuitive inferences not based on reasons, we could not say that reasons are primarily for post-hoc rationalisation. I see no grounds for Mercier and Sperber’s conclusion that the ‘main role of reasons is not to motivate or guide us in reaching conclusions but to explain and justify after the fact the conclusions we have reached’ (2017, 112). No matter how often we post-hoc rationalise our intuitive, habitual behaviours to retain a decent social standing, the defining, significant choices, those that make us the persons and indeed the species we are, are still those based on reasons. We write books, found companies, build spaceships, and play symphonies in orchestras for good reasons. We choose to do those things because we take ourselves to have good reasons for doing them.

Indeed, the discussion above shows that the question of ‘what are reasons for?’ (M&S, 2017, 123) is misleading. Reasons are facts that speak in favour of certain responses (Scanlon 2014; Broome 2004). Leaving metaethical and ontological details of that assumption aside, the significant question, when we want to understand human cognition, is the following. What do we see facts as
reasons for? In other words, why do we have the capacity to see ordinary facts about the world, like the fact that the street is wet, or the fact that someone is in pain, as reasons for us to take certain actions, or to believe certain things? The (partial) answer this article gives is: to engage in individual or group reasoning and to play the social game of giving and asking for reasons (Brandom 1982). Reason is for figuring out what we ought to believe and do, as well as for exchanging justifications with others, be it for facilitating communication, establishing trust, or managing reputations and expectations.

Thus, without downplaying the guiding function of reasons and the importance of reflective reasoning, one can accept the interactionist view that at least one function of Reason is to ‘produce and evaluate justifications and arguments in dialogue with others’ (M&S 2017, 10). However, in light of the analysis above, I believe a comprehensive account of the evolutionary function and origin of Reason requires a better understanding of the capacity to see facts as reasons. Only then can we investigate which manifestations of that capacity, and more generally which types of thinking, are adaptive and which are not. In the remaining space, I cannot offer a full account of what seeing facts as reasons consists in but the following remarks shall serve as a starting point.

To begin with, the capacity to see considerations as normative reasons is a person-level capacity that agents, not (mental or bodily) parts of agents, possess. It is an inherently normative capacity. We represent the world not only as being in a certain way but as favouring or requiring certain responses. Suppose, for example, I remember that it is my mother’s birthday tomorrow and based on this consideration decide to call her tomorrow. I not only grasp that it is her birthday but also see this fact as a reason to call her. I thereby manifest – without making explicit – my underlying values, that is, my beliefs about what is good, desirable, or what one ought to do. Maybe I believe it is my duty to call her, or that my call will make her happy and that it is good to increase happiness in the world. To take another example, suppose your friend mentions that she needs surgery she cannot afford. You understand not only the facts of the matter but also see them as a reason to offer help. You thereby manifest your commitment to solidarity or reciprocity, even if these values do not explicitly appear in your thoughts, and even if you have never explicitly considered these things to be valuable. Our (implicit) beliefs about what matters determine what we see as good or bad reasons for a certain kind of action.13 Seeing

13 The capacity to see considerations as reasons thus understood can explain why appealing to certain reasons in public discourse, or giving certain reasons as justifiers for our
things as reasons requires the general capacity for normative thought, that is, the capacity to think not only about what is the case but also about what ought to be. According to Rakoczy and Schmidt (2013, 17), humans indeed possess this ability from early on (age 2–3). They give empirical evidence for the thesis that human cognition has ‘a fundamental normative dimension’ and involves taking a normative, agent-neutral stance to how one does things, or to how things are done.\footnote{All of this allows for cultural differences: \textit{which} reasons we see as how weighty depends on our social setting and culture (O’Madagain 2019).} On a more philosophical note, the capacity to recognise reasons arguably explains the status of free persons who are responsible for their attitudes and actions (Pettit and Smith 1996; Scanlon 1998; Raz 2011; Wallace 2006; Pettit 2007).

If the capacity to see things as reasons is central to Reason and requires the capacity for normative thought, then the evolutionary origin of Reason is linked to the evolutionary origin of this capacity. While I cannot pursue this topic in detail here, the capacity to see things as reasons appears to be conceptually very close to what Tomasello (2016, 82) calls ‘the human sense of ‘ought’.

According to Tomasello, this mental capacity developed in early humans when they moved from joint intentionality to collective intentionality in response to the novel need for cultural collaboration (Tomasello 2016, 82):

\begin{quote}
Early humans’ cooperative rationality expands human pro-attitudes to include the welfare of others, it presupposes second-personal agents who consider one another as equally deserving of respect and resources and it focuses on the individual decision making that takes place within the context of the joint agent, “we”, formed by a joint commitment. These new elements in the decision making of individuals created a socially normative dynamic force behind their actions.
\end{quote}

The sense of ‘ought’, the idea that there are objectively correct ways of doing things, is ‘an adaption to problems of social coordination in collaborative foraging’ (ibid, 79). If my hypothesis is correct and this sense of ‘ought’ underlies actions, commits us to give similar reasons in the future (M&S 127, 186): When giving reasons, we express, (without making explicit) the fundamental values we accept. Precisely because we are not expected to change those value commitments frequently, people can expect that those values will guide us in the future, too. They can expect that these values will continue to determine what we see as a good or bad reason, which allows them to manage their expectations of our actions accordingly (see also Stanford, Thomas, and Sarnecka’s (2020) reply to Cushman (2020)). This result of my analysis supports Strijbos and de Bruijn’s (2015) account of confabulating rationalisations as shaping our value judgments.
our capacity to see features of the world as normative reasons for certain responses, then the approach to Reason I proposed points towards a social, interactionist account of the evolution of Reason that gives special weight to in-group cooperation. This, in turn, suggests that Reason is a group-level adaptation to cooperation needs (Tomasello 2014) rather than an individual-level adaptation to in-group competition for reputation (M&S 2017) or to the need for efficient individual reasoning (Cushman 2020). Importantly, the claim that Reason, and thus reasoning, has a social origin is compatible with the claim that reasons (or better: our capacities to recognise them) enable not only the social game of giving and asking for reasons but also individual decisions and judgments (O’Madagain and Tomasello 2019). The view is, to borrow a phrase from Nagel (1996, 202), that ‘the reflective self [the one who thinks whether impulses comply with reasons] is in its nature more universal than the original, unreflective self because it achieves its self-conscious awareness by detaching from the individual perspective’.

5 Conclusion

I defended the common-sense paradigm that we are guided by reasons in thought and action against central arguments of what I called the Rationalisation View. First, I argued that we should uphold the distinction between, on the one hand, reflective, reason-guided reasoning and, on the other hand, intuitive, automatic thought processes. The former essentially involves the capacity to see things as reasons. Second, I investigated post-hoc rationalisation and self-serving rationalisation and found that both involve the capacity to see things as reasons, too. This finding led me to the conjecture that the capacity to see considerations as (normative) reasons is essential to and a core feature of Reason, which cannot plausibly be understood as a module or mental mechanism. Reason, on this view, is manifest in reflective reasoning, as well as in rationalisation. While this supports the common-sense paradigm, it also implies that some instances of rationalisation are indeed rational in a defined sense. Third, regarding the function and origin of Reason, I argued that we should ask: why do we have the capacity to see facts as reasons? While I could not provide a full account of what seeing things as reasons consists in here, the hypothesis that it is closely related to that which Tomasello calls the human sense of ‘ought’ led to the thesis that Reason is inherently social and a group-level adaptation to collaboration needs. Investigating this topic further seems important, not least because it shapes our perception of ourselves. Should we accept biases, automatic rationalisations and lazy patterns of thought as welcome features
of our minds that serve the individualistic function of Reason, or should we think of them as residues of a distant past that we ought to overcome through better use of Reason?

Acknowledgments

For helpful discussions on this article, I thank Anna Strasser, Johannes Brandl, Hans-Johann Glock and the research group of the project “The structure and development of understanding actions and reasons”. For financial support, I thank the Swiss National Science Foundation (SNF, #5100019E_177630).

References


