Why did you really do it? Examining the distinction between kinds of reasons

José Ángel Gascón
Universidad Católica de la Santísima Concepción

Follow this and additional works at: https://scholar.uwindsor.ca/ossaarchive

Part of the Philosophy Commons

Gascón, José Ángel, "Why did you really do it? Examining the distinction between kinds of reasons" (2020). OSSA Conference Archive. 1.
https://scholar.uwindsor.ca/ossaarchive/OSSA12/Wednesday/1

This Paper is brought to you for free and open access by the Conferences and Conference Proceedings at Scholarship at UWindsor. It has been accepted for inclusion in OSSA Conference Archive by an authorized conference organizer of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.
Why Did You Really Do It? Examining the Distinction Between Kinds of Reasons

JOSÉ ÁNGEL GASCÓN
Centro de Investigación en Educación y Desarrollo (CIEDE)
Universidad Católica de la Santísima Concepción (UCSC)
Alonso de Ribera 2850, Concepción
Chile
jgascon@ucsc.cl

Abstract: Studies in cognitive psychology have shown that many of our actions seem to be often influenced by irrelevant features of the environment, of which we are not aware. But exactly what reasons has the psychological research uncovered? In philosophy, a distinction has been made between normative, motivating, and explanatory reasons. Hence it is necessary to determine which of them have been revealed as the real reasons for our actions by the psychological research.

Keywords: Cognitive psychology, explanation, justification, motivation, rationalisation, reasons for action

1. Introduction

Human beings consider, at least sometimes, what reasons we have to do something. When we do, we usually act in the light of those reasons that seem to us to be the best. Imagine, for example, that I have been offered two jobs, one of which has a better salary and the other is in a city that I like. Which job I will accept depends on my weighing of those reasons and of which considerations are more important to me. After my decision, if I am challenged to justify it—“why did you do that?”—I will present those reasons that made me opt for one job rather than the other, hoping that the other people will see why I made the best choice. I will, then, engage in argumentation in order to show that those reasons that moved me to act in a certain way were the best reasons, all things considered.

This can be considered as the standard, common-sense view of reasoned action and justification of actions. We justify our actions by presenting reasons, and those are precisely the reasons for which we acted. Characterisations of the rational person or the critical thinker which focus on reasons—as opposed to those which focus on the suitability of means to an end, for instance—tend to rely on this view of the relationship between reasons and action. According to Harvey Siegel, for example, a critical thinker is someone who is “appropriately moved by reasons” (1997, p. 49). And, in the literature on normativity and practical reasons, authors such as Scanlon (2014) and Kiesewetter (2017) define rationality in terms of responsiveness to reasons.

The idea that we should justify our actions by putting forward the reasons for which we acted seems like a plausible one. After all, there is a tendency to see people as irrational—or, at the very least, hypocritical—when they act for one reason and afterwards attempt to justify their action by appealing to different reasons. Consider the case of someone who decides to study philosophy and holds that her reason for that decision is her love of knowledge, when in fact what moved her towards a philosophical career is her desire to enjoy a high cultural status. No doubt many of us would see that behaviour as falling short of rationality—or, if she is aware of her real motivation, as insincere.

Indeed, the most common use of the term “rationalisation” refers to cases of that sort. A rationalisation, in this negative sense, is an attempt to explain an action performed by oneself that fails because it does not identify the considerations in light of which one acted in that way. As Audi explains (1985, pp. 159-160):
Many purported explanations of action fail because they are merely rationalizations, as where a person cites an altruistic reason he had for helping someone, when in fact he was motivated by selfish reasons. sort of case is the most typical kind of rationalization.

People usually dismiss as “mere rationalisations” the reasons that they are offered when they suspect that those reasons do not explain the action. Such reasons, it is often said, would not be the “real reasons”. When people ask for reasons for an action, they usually want the reasons for which the action was performed, and it is those reasons—if anything—that can justify the action.

However, if this is how we should understand the rational justification of actions, then apparently we are in serious trouble. The empirical research in the psychology of reasoning has shown that human beings are very bad at identifying the causes of our own actions. A growing number of empirical studies have provided evidence that we lack access to the knowledge of what considerations move us when we act. The reasons that we put forward when we are challenged to justify our behaviour are not, it seems, those reasons for which we acted; they are merely our best guesses about what could make our actions plausible—even though no doubt those guesses are sometimes right. Thus, in that sense, most of our justifications of actions seem to be rationalisations. As Mercier and Sperber put it (2017, p. 253): “Humans are rationalization machines.”

How worried should we be by this conclusion? The purpose of this article is to give a tentative answer to this question. I believe that any such answer, if it is to be plausible, must be both philosophically and psychologically informed. Our philosophical accounts of practical reasoning need to take into account the empirical findings that indicate what feats human reason can and cannot achieve; and, at the same time, the psychological research must be based on a philosophical understanding of reasons so that it is clear what conclusions can and cannot be drawn from the empirical data. I will begin, in the next section, by reviewing the empirical studies in psychology of reasoning that cast doubt on our ability to detect the reasons that move us to act. Then, in section 3, I will present philosophical distinctions between kinds of reasons and I will provide an interpretation of the conclusions of psychological studies in the light of those distinctions. Finally, in section 4, I will draw some preliminary conclusions about how all this should affect our conceptions of justification of actions and of rationalisation.

2. Psychological research on reasons for action

Up until the 1970s, it was widely assumed by psychological researchers that we are aware of the mental processes that lead to our judgements and our behaviour (Kunda, 1999, p. 265). In order to study people’s choices and evaluations, investigators resorted to self-report questionnaires in which the participants in the experiments were asked to state why they behaved as they did. Researchers who attempted to study the grounds for liking a political candidate or for choosing a job, for example, simply asked people why they liked a certain candidate or why they chose a certain job. Even behaviour in hypothetical scenarios was sometimes studied by asking people what they would do in the situation in question. However, it eventually became manifest that such self-reports are not reliable.

In their ground-breaking article, Nisbett and Wilson (1977) reviewed a series of empirical studies in which a particular stimulus demonstrably influenced the participants’ actions and judgements but, when interviewed, the participants denied that influence and tended to explain their behaviour by reference to other factors. An example is the large number of experiments that showed the existence of the “bystander effect,” the fact that
people are less likely to help a person in distress if there are many other onlookers around (Latané & Darley, 1970). After the experiments, Latané and Darley asked the participants whether their decision to help or to abstain from helping had been influenced by the presence of other people. Despite the robust evidence that showed that a greater number of onlookers correlated with a failure to help, the participants systematically denied that influence. As the authors explain (p. 124):

We asked this question every way we knew how: subtly, directly, tactfully, bluntly. Always we got the same answer. Subjects persistently claimed that their behavior was not influenced by the other people present.

Nisbett and Wilson conducted a series of small studies in order to investigate the accuracy of causal explanations of one’s own behaviour (Nisbett & Wilson, 1977; Wilson & Nisbett, 1978). The experiments were designed in a way that resembled as close as possible situations of the real life, with little or no deception involved. Yet they were also designed so that the stimuli that would probably influence the participants’ behaviour were of a counter-intuitive sort and hence their influence could not be accounted for by the participants’ prior causal theories of how people behave (Nisbett & Wilson, 1977, p. 242). Therefore, those stimuli could only have been detected by the participants if they had genuine introspective access to their own cognitive processes. As expected, people were influenced by factors whose influence they could not detect—and, interestingly, the researchers themselves were highly unsuccessful in their predictions of which factors would influence them.

In one of those studies (p. 243), the participants were asked to evaluate four pairs of stockings. They had to chose one of those pairs and, afterwards, they were asked why they had chosen it. The trick was that all the stockings were identical. Nisbett and Wilson observed that the stocking situated towards the right were preferred over the ones situated at the left. However, when the participants were asked about the reasons for their choices, the position of the article was never mentioned. In fact, when the researchers suggested that possibility to the participants, they denied it. The authors explain that (Wilson & Nisbett, 1978, p. 124):

Only a quarter of the subjects required any prompting to explain the basis of their choices. Most of the subjects promptly responded that it was the knit, weave, sheerness, elasticity, or workmanship that they felt to be superior. […] Not a single subject mentioned the position of the stockings as a reason for the choice.

Not only do we often fail to detect factors that cause our behaviours, but we also tend to report as reasons for our choices and judgements stimuli that actually had no effect on us. For example, in another experiment (Nisbett & Wilson, 1977, p. 246), the participants had to predict how much electric shock they would take. Some of them were said that the shocks would do “no permanent damage,” while the others were not given that reassurance. Then, the researchers asked the first group whether that comment had affected their predictions, and they asked the second group whether, had they made that comment, their predictions would have been different. Inclusion of the reassurance proved to have no effect on the predictions of how much shock the participants would take, but a majority of them reported that it affected their predictions.

What all this evidence means is not merely that we are sometimes wrong when we report our reasons for our decisions and judgements—that would hardly be big news. Neither can it be concluded that we are always wrong; as Nisbett and Ross (1980, p. 211) admit, we are often accurate in our explanations of the reasons for our behaviour. The worrying
implication of that research on self-reports is rather that we lack introspective access to the reasons that guide our behaviour. The process by which we arrive at a belief of why we did something is the same whether that belief is accurate or inaccurate: we infer it from the known data and from our prior theories of human behaviour. That is, it is the same process that we follow when we propose causal explanations of other people's behaviour (Ibid.). If, for example, I buy a bottle of water and I claim that I did so because I was thirsty, I am surely right. But this is so simply because we have a common-sense theory of why people usually buy bottles of water, and that theory is largely correct. Notice, also, that it would be just as easy to identify the reason why someone else bought a bottle of water. The problems begin when there is no prior theory or when that theory does not fit the case at hand. If we fail to help a person in distress because there are many other people around, or if we choose a pair of stockings because they are situated on the right, then we are likely to give a wrong account of our behaviour, since we have no prior theory about the relationship between those reasons and those actions. And, in those cases, we are just as likely to be wrong about our own behaviour as we are to be wrong about other people’s behaviour. The process is the same in both cases.¹

If that conclusion is correct, then the processes that cause our actions and judgements are unconscious, just as the processes that are responsible for perception or textual comprehension. Kunda (1999, pp. 270ff.) reviews other studies that provide evidence of those unconscious processes that influence our behaviour, including aspects such as implicit memory, subliminal perception and affect. Even though the conclusion that we never have introspective access of the stimuli that cause our actions is still controversial, some cognitive scientists have accepted its most dramatic implications regarding our conscious will. Evans, while admitting that there is a difference between voluntary and involuntary actions, questions the very existence of a conscious will (2010, p. 177):

‘We’ are not conscious persons in control of our behaviour and the reflective mind does not equal a conscious mind. The conscious person is a construction of the brain, an illusory narrative that accompanies us through life.

In the same vein, Wegner (2002) talks about the conscious will as an “illusion.” According to his theory of apparent mental causation, conscious will is not a cause of actions but simply a (possibly misguided) feeling that an action was caused by us. He explains (p. 336):

Apparent mental causation suggests that the experience of consciously willing an act is merely a humble estimate of the causal efficacy of the person’s thoughts in producing the action. Conscious will is the mind’s way of signaling that it might have been involved in causing the action. The person’s experience of doing the act is only one source of evidence regarding the actual force of the person’s will in causing the action, however, and it may not even be the best source.

Now, if we accept these psychologists’ conclusion that we are never introspectively aware of what factors influence our actions and judgements, and in fact we are often wrong about them, the question is: how big a problem is that for our philosophical theories about reasons for action and justification of actions? This is a very broad issue that cannot be solved in a single paper. As a first step, however, it would help to be clear about what exactly Nisbett and Wilson’s experiments uncovered. Did they identify our real reasons for action? Or did

¹ Even though, of course, when it comes to our own behaviour we have access to data that we lack when we attempt to interpret someone else’s behaviour, such as feelings, goals, beliefs or memories (Nisbett & Ross, 1980, p. 203).
they show us simply the causes of our actions? Are they the same thing? Sorting out this conceptual issue is the purpose of this paper, and to this I move in the next section.

3. What reasons are we talking about?

The results of the experiments conducted by Nisbett and Wilson certainly seem to reveal something important that jeopardise our ideas of intentional action and justification of actions. But, what is it exactly that was identified in those experiments? In their articles, Nisbett and Wilson used a variety of terms to refer to that finding: “influences,” “explanation,” “causes,” “causal factors,” and “reasons” for choice. The point was that there seemed to be a mismatch between the reasons stated by the participants in the studies and whatever it was—influences, reasons, causes—that explained their choices. Thus, a necessary first step in the assessment of the implications of those studies for our philosophical theories is the clarification of these factors that explained the participants’ behaviour.

The most natural interpretation, I believe, is that the experiments identified the causes of our actions and judgements. Now, it is well known that, according to certain philosophical views, reasons for action just are the causes of our actions. Davidson (1963) famously argued for that view. If that is how we should understand practical reasons, then the discovery that people lack direct awareness to the causes of their actions obviously challenges our practice of giving reasons for our actions. If we cannot detect the causes of our behaviour, and practical reasons are precisely those causes, then it seems that our justifications of actions are mere fictions.

However, the philosophical literature has distinguished between different kinds of reasons, and Davidson focused on only one of them: the kind of reason that “explains the action by giving the agent’s reason for doing what he did” (p. 685). The same can be said of Searle (2001), who clearly differentiated between justification and explanation of actions, and stated that the latter must be a causal account (p. 110, his emphasis):

…the explanation of why something should have been done or is a good thing to have been done is not always the same as why it was in fact done. In this book we are primarily concerned with explanations that explain why something happened, with explanations that state the reasons that the agent acted on or will act on. We are interested in justifications only insofar as they also explain why the agent acted or will act. Therefore I will distinguish between justifications and what I will call “justificatory explanations.” Justification does not always explain why something in fact happened, but an explanation of its happening, whether justificatory or not, has to explain why it happened.

Thus, it is one thing to report what considerations motivated us to do something, and hence explain our action; it is something different to justify our actions with considerations that make them the right thing to do (cf. Dancy, 2000, pp. 20–25). The former kind of reasons has been called motivating reasons, whereas the latter has been called normative reasons. Thus, Parfit (1997, p. 99) says that normative reasons are those that we are looking for when we ask “What do we have most reason to want, and do?”; motivating reasons, on the other hand, are those in the light of which we act. Dancy explains the distinction this way (2000, p. 2):

There is the question what were the considerations in the light of which, or despite which, he acted as he did. This issue about his reasons for doing it is a matter of motivation. There is also the question whether there was good reason to act in that way, as we say, any reason for doing it at all, one perhaps that made it
sensible in the circumstances, morally required, or in some other way to be recommended, or whether there was more reason not to do it…. This second question raises a normative issue.

Thus, we can act for a good reason, in which case our motivating reason is also our normative reason, but it is also possible for these two kinds of reasons to diverge. Imagine, for example, that I voted for a certain political candidate because it seemed to me that she was the most honest and competent one. Those were the reasons that I considered when I was deciding my vote, so those are the reasons for which I acted. When asked, I offer those reasons to justify my choice. In this case, my normative reasons are the same as my motivating reasons. But let us imagine a slightly different scenario. Imagine that, even though that political candidate was indeed the most honest and competent one, I did not take that fact into account when deciding my vote; instead, what motivated me to vote for her was that she was born in the same city as me. I still justify my vote before others by mentioning her honesty and competence, but I know that I voted for her because we were born in the same place. In this second case, my motivating reasons are different from my normative reasons.

Let us differentiate, then, between:

- **Normative reasons**: Considerations that make an action the right thing to do, that count in favour of doing that action.
- **Motivating reasons**: Considerations that moved me to do something, those in the light of which I acted.

As Dancy (2000, p. 2) and Alvarez (2009) argue, this reference to two “kinds” of reasons should not be understood as implying that there are really two sorts of reasons—reasons that motivate and reasons that justify. They are different kinds of reasons only in the sense that they are offered in answer to two different questions: (1) what makes that action right?, and (2) why did you do that action? Now, if we go back to Nisbett and Wilson’s experiments and ask what kind of reasons—if any—they have discovered, it is clear that we can rule out normative reasons. The purpose of the psychological experiments was not to find out what reasons people give to justify their actions. In fact, the reasons that the participants provided in the studies were put into question by the researchers—that was the point. The purpose was rather to check the accuracy of people’s reports of their reasons for action. The focus was on explanation of actions, not on justification.

However, it is not clear to me either that the findings of the experiments refer to motivating reasons. Those findings do refer to factors that explain people’s actions, but motivating reasons are not simply any kind of explanation; motivating reasons explain actions only insofar as those actions were made *in the light of* reasons. That means that a causal factor would not count as a motivating reason if the agent has not consciously considered it and decided to act on the basis of it. For a cause of people’s actions to be a motivating reason, they must at least recognise it as a reason and be guided by it. Suppose, for instance, that I fell on the street because a car hit me. It does not make sense to say that my motivating reason for falling was that a car hit me, i.e. that I was motivated to fall by the hit of a car. No doubt the hitting of the car explains the event, but it is not an explanation in terms of *motivation*.

The relationship between the factors that influenced the participants in Nisbett and Wilson’s experiments and the participants’ choices seems to be of a similar kind as the relationship between the hitting of the car and my fall. As we saw, in the experiment of the stockings, none of the participants cited the position of the stockings—the causal factor—as a
reason for their choice. Similarly, in Latané and Darley’s experiments, the participants denied that the number of people present influenced their decision to help someone in distress. Thus, if those factors explain the participants’ behaviour, the explanation cannot involve motivation through reasons.

There is, however, a third possibility besides normative and motivating reasons: what Searle (2001, p. 111) calls “straight causal explanations” and Dancy (2000, p. 5) calls “reasons why.” These reasons do not involve considerations that the agent takes to favour some action. Actions that are explained on the basis of “reasons why” are not performed in light of those reasons, but simply caused by them. This is the case with the explanation of my falling by citing the hit of the car. Many other events involve this kind of explanations, in which no reason was considered by the agent, as Dancy reminds us (Ibid.):

What explains why one person yawned may be that someone else yawned just next to them. What explains why he responded so aggressively may be that he is having trouble at home or that he has taken a particular form of medication. What explains why he gave this student a better grade than she deserved is that he was unconsciously influenced by the fact that she always dresses so neatly (or something even less defensible). What explains why so many people buy expensive perfume at Christmas is the barrage of advertising on the television. What explains why he didn’t come to the party is that he is shy. In none of these cases are we specifying considerations in the light of which these things were done.

Dancy states that what these explanations involve “is not a reason at all, really, but rather a cause” (p. 6). But Alvarez (2009, p. 184) argues that its being a cause does not exclude its being a reason, since both terms belong to different domains: that of causation and that of explanation. We use reasons to explain actions, and those reasons sometimes happen to be causes in the natural realm. Therefore, she proposes that, besides normative and motivating reasons, we should consider explanatory reasons. If we differentiate among different kinds of reasons on the basis of the role they play in answering different questions, then the question of what explains an action is substantially different from the question of what motivated the agent—even though, of course, the same reason can answer both questions.

We have, then, according to Alvarez’s proposal, three kinds of reasons:

- Normative reasons: Considerations that make an action the right thing to do, that count in favour of doing that action.
- Motivating reasons: Considerations that moved me to do something, those in the light of which I acted.
- Explanatory reasons: Considerations that explain why I did something, what caused my action.

Now, explanatory reasons are a better candidate for the kind of reasons that the psychological experiments revealed. They are causes that explain people’s actions without being at the same time motivating reasons, since people do not consider them and even deny their influence. They are causes in the same sense that taking a certain medication is the cause of aggressive behaviour. They influence our actions but we are unaware of that influence.

There is one crucial difference between explanatory reasons and the other two kinds of reasons, and that difference is what makes the findings of psychological experiments so shocking: explanatory reasons do not necessarily involve human agency. Just as they can be
used to explain human actions, they are also what explains events such as the rain, the collapse of a building or the movement of waves at sea. There are no normative or motivating reasons for events like these—water and buildings do not consider reasons and do not attempt to justify their actions. So, when human actions are explained on the basis of explanatory reasons that are not also normative or motivating, that certainly feels like our sense of agency itself is being challenged. That may be all right for certain human actions, such as yawns or sudden outbursts of aggressiveness, but it is frightening to find out that it also involves actions for which we believe we have motivating reasons, such as choosing stockings of helping a person in distress. No wonder some cognitive scientists have concluded that conscious will is an illusion.

So, should we give up any talk of normative or motivating reasons altogether? My answer would be that, once we have distinguished among kinds of reasons and have specified the role each of them plays, we can resist that conclusion. That is why it was so important to be clear about what kind of reasons we were talking about, and why that is the main topic of the present article. Of course, satisfactorily solving this issue would require a whole paper in itself, but I will attempt to outline at least the beginning of an answer. That is a task for the next section.

4. Justification, motivation and rationalisation

The practices of considering reasons for action and of exchanging reasons are specifically human. Our consideration of reasons allows us to say that our actions are not merely things that happen to us, but things that we wilfully do. Our ability to explain and justify our actions distinguishes us from dogs, trees and thermostats. But if the reasons that we report have no influence on our behaviour, as the experiments that we have seen suggest, then it begins to look as if reasons were merely epiphenomenal: they would play no role in the determination of our actions.

Before drawing that conclusion, however, we should examine more closely a kind of reasons that, as we saw in the previous section, the psychological experiments did not address: normative reasons. Even though we often attempt to justify our actions by explaining why we performed them—by citing motivating reasons—normative reasons need not be also motivating reasons (Dancy, 2000, p. 3). Sometimes we simply argue that what we did was right without intending to explain what moved us to do it. Someone might, for example, argue that the choice of her academic career was a good one—because, say, it had good job prospects and it fitted her character—without even remembering why she chose it in the first place. Furthermore, normative reasons can refer to future actions as well as past actions, they arise when deliberating about what to do, and in that case there is no point in asking if those reasons are also what motivated us to do it—nothing has been done yet. So, arguably, the reality of normative reasons is not affected by discoveries in the empirical sciences about what influences our behaviour. That is just not their main role.

Nevertheless, normative reasons would be rather pointless if they did not have any effect on behaviour at all. But I believe that both our everyday experience and the empirical research provides grounds for concluding that, at least sometimes, normative reasons do influence decisions. The idea that people can take decisions and change their minds on the basis of reasons that show that some action is the right thing to do seems to be a necessary assumption in order to account for much of human behaviour. This can be seen most clearly in psychological experiments involving interpersonal argumentation. As Mercier and Sperber (2017, pp. 264–265) point out, groups of people are more able to solve logical problems than individuals working alone, and this happens because people working in groups get convinced by good reasons. For example, Trouche, Sander and Mercier (2014) showed that people who
are confronted with arguments or who have to argue are more likely to solve logical problems such as those of the Cognitive Reflection Test (Frederick, 2005) and others. The experiments were designed in a way that ruled out the effect of degrees of confidence of some participants on others, measuring specifically the effects of good argumentation. Thus, they concluded that their results “make it clear that arguments, rather than confidence, are the main factor explaining the performance of groups discussing intellective tasks” (p. 1968).

According to Mercier and Sperber, the human faculty of reason does not lead us to make better decisions but—at least sometimes—to make decisions for which we can come up with good reasons (p. 255): “when people have weak or conflicting intuitions, reason drives them toward the decision for which it is easiest to find reasons—the decisions that they can best justify.” In such cases, then, the reasons that explain our behaviour might not be the reasons that we think—they are not those that support the best decision, as opposed to the most justifiable decision—but they are normative reasons nonetheless. So it seems that normative reasons can play a role in determining our actions. Normative reasons can influence us in group discussion, as Trouche et al. showed; or, even when there is no interpersonal argumentation taking place, the potential justification that we mentally rehearse leads us in the direction of the most acceptable normative reasons.

Hence, normative reasons do not seem to be inert. They can lead people to take a decision or form a judgement. But then, when such a thing happens, we can confidently say that those are people’s motivating reasons. Just as in Nisbett and Wilson’s experiments there was no way that alleged motivating reasons played any role in determining behaviour, in the experiments reviewed by Mercier and Sperber there seems to be no alternative to granting normative reasons a causal role. Therefore, reasons are not a mere epiphenomenon; motivating reasons, i.e. conscious reasons with a causal power on our decisions, exist.

Concluding that motivating reasons exist, however, is not concluding much if those reasons can only be reliably detected in the laboratory. And that is precisely the lesson that we should draw from the psychological research: people do not have direct access to the causes of our own actions, we just infer the possible causes from a body of data and a more or less accurate theory of human behaviour, so we can always be wrong about our alleged motivating reasons. We should not be confident that we did something for the reasons that we think we did it. We need to accept our unreliability even in the realm of our own actions. That does not mean that we are always wrong, of course. It simply means that there is no privileged position from which reasons for action can be detected, not even the position of the agents themselves. This is Nisbett and Ross’s conclusion (1980, p. 211):

Empirically, this means that under most circumstances subjects will be right in their causal accounts if and only if observers, working with similar externally available information, also are right.

Taking this conclusion seriously should, in my view, lead us to giving considerably less weight to motivating and explanatory reasons in people’s attempts to justify—or criticise—actions. Accounts of why we did something, for which reasons we did it, should not be given a predominant place in justifications of actions. Normative reasons should not depend, and should be kept separate from, motivating reasons. I believe this is a conclusion that we have to consider seriously in light of the unreliability of our reports of motivating reasons. If we do not want the weakness of those reports to be transferred to our practice of justifying actions, the kind of reasons that make an action right or wrong should be relatively independent of the kind of reasons that explain why that action was done.

This may seem too radical a proposal, as it blurs the distinction between a genuine justification of an action and a rationalisation. Without that distinction, it may be thought that
the very idea of rationality is in danger. I will use the rest of this section to attempt to dispel that worry.

In the most common sense of the term, a rationalisation is a purported account offered by an agent of one of her actions that (Audi, 1985, p. 163):

1. Offers one or more reasons for doing that action.
2. Represents his doing that action as at least *prima facie* rational given those reasons.
3. Does not explain why the agent did that action.

That is, a rationalisation is an attempt to *justify* an action (point 2) by offering *normative* reasons (point 1) that are not at the same time *motivating* reasons (point 3). Someone might, for example, justify his decision not to eat peppers by asserting that they are bad for his health, when in fact his motivating reasons is simply that he does not like them. But rationalisation so defined is exactly what, I have argued, theories of rational action should be more tolerant of. We should accept the fact that, as Mercier and Sperber (2017, p. 253) put it, humans are rationalisation machines.

Is this a call for widespread irrationality? I believe that we can dissipate that fear if we do not underestimate the extent to which normative reasons can be criticised. A match between normative reasons and motivating reasons is not the only way to check the correctness of justifications—it is not, in fact, the main one or even the most demanding one. Normative reasons by themselves must fulfil several criteria for them to constitute a satisfactory justification. One of those criteria is, of course, that normative reasons must be true, they must mention real facts. This criterion already allows us to see where the participants in Nisbett and Wilson’s stockings experiment got their justification wrong: they mentioned particular features of the stockings that they chose that made them the best stockings in the lot, whereas in fact all the stockings where identical. There is no need to appeal to their motivating reasons in order to conclude that their justifications were flawed.

Besides truth, we should expect an agent’s normative reasons to cohere with the normative reasons that the same agent has offered in similar circumstances. The principle that like cases should be treated alike is firmly established both in law and in ethics. The normative reasons offered in a particular case, therefore, should not be assessed in isolation from those offered in all the other cases that the agent has encountered. This principle helps us explain what might be wrong in the justifications offered by the participants in the bystander effect experiments performed by Latané and Darley. Surely, those who helped the person in distress in the scenario with few onlookers could have justified their action by saying that the person needed help, but if they do not help in a similar scenario with more onlookers, there might be an incoherence in the normative reasons they state.² Again, as in Nisbett and Wilson’s experiments, we can talk about the rationality or irrationality of justifications without looking for motivating reasons.

Consider, finally, Audi’s example, mentioned in the Introduction (Audi, 1985, pp. 159–160): “a person cites an altruistic reason he had for helping someone, when in fact he was motivated by selfish reasons.” If what explains that action is selfish reasons, one would expect that the person would not behave the same way in a situation in which she again must help someone but the selfish reasons are absent—there is no benefit for her. That would reveal an incoherence in her attitude towards normative reasons between both cases.

² I say that there *might* be an incoherence because I am not sure that there is no relevant difference between the two scenarios to which the agent could rightly point out. After all, if there are many onlookers, the agent could always argue that she thought that someone would take care of the person in distress, and perhaps that is a legitimate expectation.
Otherwise, if her behaviour was consistently helpful, insisting on her selfish motivation in order to criticise her normative reasons would seem to entail a moral theory that is way too demanding.

All this is not intended to mean that we should never take into account motivating reasons when assessing justifications. Even within the boundaries of a single action, a clear mismatch between normative and motivating reasons can be reprehensible. If it is clearly apparent, for example, that I intended to punch someone out of anger and, by sheer luck, I ended up moving him away from a bus that was going to run over him, saving his life, then I can hardly justify my action by saying that I saved his life. Anyone could see that my intention was to hit him. However, apart from clear cases like this one, our practice of giving and asking for justifications should not focus on mismatches between normative, motivating and explanatory reasons. We should accept that those mismatches are ubiquitous in human action, as the research in experimental psychology has shown, but at the same time we can be confident that we have the resources to assess normative reasons by themselves.

5. Conclusion

Research in cognitive psychology during the last five decades has shown that, in many situations, the reasons with which people explain their own actions and judgements do not correspond to the real factors that caused them. This finding has led to the conclusion that people do not have introspective access to the causes of their own behaviour; instead, people infer them, just as they would if they were observers of someone else’s behaviour. Such a conclusion seems to cast doubt on the significance of our practice of justification of actions and exchange of reasons. However, in order to fully understand the philosophical implications of the results of psychological research, we need to be clear about what kinds of reasons are psychologists talking about.

In the philosophical literature, three kinds of reasons have been distinguished, according to the kind of question that they answer: normative reasons, considerations that make an action right; motivating reasons, considerations in light of which the person acted; and explanatory reasons, considerations that explain what caused an action or an event. The problem with the psychological experiments, we saw, was that the participants offered purported motivating reasons that did not explain their choices at all; instead, what explained their choices was explanatory reasons that the experiments uncovered and of which the participants were unaware. That challenges the reality of motivating reasons, and we are left only with normative reasons that, for all we know, could have no effect on behaviour whatsoever—they could be epiphenomenal.

However, we also saw that certain behaviours could only be plausibly accounted for by the influence of normative reasons. If our performance in a logical task is better when there is argumentation, and if we tend to lean towards the most justifiable decisions when our intuitions are weak, that gives us grounds for believing that sometimes normative reasons do guide our actions. So normative reasons can also be motivating reasons. The problem, given our lack of introspective access to the causes of our behaviour, is that in practice we can never be sure that, in a particular instance, we are genuinely motivated by normative reasons. For this reason, I argued that our assessments of the normative reasons provided by agents should not give much weight to whether they are also motivating reasons or not—i.e., whether they are rationalisations of actions. Outside laboratory conditions, the identification of motivating reasons is a tricky issue and it is bound to lead to speculations, and we have the conceptual resources to assess normative reasons in themselves.
Acknowledgements: This research was possible thanks to the postdoctoral scholarship CONICYT/FONDECYT n. 3190149.

References