

The desire for parsimony – to posit as few explanatory features as possible – has a rich philosophical history and is often given lots of weight in philosophical theory construction. But, as the psychologist Tania Lombrozo has demonstrated, our bias in favour of parsimony can lead us to adopt simple explanations even when it's far more likely that a complicated explanation is correct.

Just what are your intentions?

MARK PHELAN ARGUES THAT OUR JUDGEMENTS ABOUT THE INTENTIONS OF OTHERS ARE NOT AS SIMPLE AS THEY SEEM

Jules is the CEO of Globotron Megacorp. One day her vice president comes to her and says, "We're ready to go ahead with the new program. The project will increase profits. Problem is, it will also *harm* the environment. What should we do?" Sitting behind her huge mahogany desk, Jules clips the end off a Cohiba and says, "I don't care about the environment. I just want profits. Go ahead with the project." At the same time, across town, Jill, the CEO of Halibrex Amalgamated, is

approached by her VP, who tells her, "We're ready to go ahead with the new project. It will increase profit, but it will also *help* the environment. What should we do?" Jill pours herself a scotch and says, "I don't care about the environment. I just want profits. Start the project." Did Jules harm the environment on purpose? Did Jill help it intentionally? Most people think that Jules intentionally harmed the environment, but that Jill did not intentionally help it. But how could this be when both express the same single-minded desire for profits and careless attitude towards the natural world?

Evaluations of people's reactions to these cases marked an important early example of

Mark Phelan is postdoctoral fellow and lecturer in philosophy and cognitive science at Yale University



experimental philosophy. Neither CEO particularly wanted to harm *or* help the environment. In each case, environmental effects were a mere *side-effect* of the desired goal of more profits. So the common opinion, that the harming CEO but not the helping one acted intentionally, challenged the dominant, philosophical opinion that deciding whether someone acted intentionally had to do only with drawing conclusions about someone's beliefs and desires. Here people seemed to be basing their intentionality judgements not on the CEOs' attitudes (identical in both cases), but rather on some aspect to do with the morality of harming versus helping the environment itself.

Since the introduction of the CEO cases, different explanations of why and when side-effects are thought to be intentional have appeared at a rate far outstripping a cottage industry (and more akin to the modern mechanised assembly line). The problem is, no

sooner does a philosopher articulate a theory of intentional side-effects than some young experimentalist blows it up by devising a new case that doesn't conform to the pattern. (A large proportion of current experimental philosophers cut their teeth devising or refuting different explanations of intentional side-effects.) I'll review a few of these explanations, and some of the cases they couldn't account for. Then I'll reflect on what this crowded history suggests about explanatory practice in experimental philosophy.

Early explanations of intentional side-effects all emphasised the morality of harming versus helping the environment. One theorist suggested that people rationally conclude that all foreseen bad side-effects are intentional. Jules foresaw that her actions would result in harm to the environment, so she harmed the environment on purpose. Another thinker claimed that people want to blame Jules for hurting the environment, but they know one can

»»

only be blamed for what one does on purpose. So, to justify their blame, they decide after the fact that Jules acted on purpose.

But neither of these explanations accommodates our intuitive opinions about another CEO, Hal. When his VP explains that a new plan will result in a decrease of sales in New Jersey, but increased profits overall, Hal says, "I don't care about the rate of sales in New Jersey. I'm just interested in the bottom line. Go ahead with the plan." Most people don't think it's bad for a CEO like Hal to decrease sales in New Jersey. Nor do they think he deserves either praise or blame for decreasing sales there. Still, despite his clear avowal of indifference, most people think Hal decreased sales in New Jersey on purpose. This is an intentional side-effect without badness or blameworthiness, so explanations in terms of badness and blameworthiness fail to account for this case.

Many second-wave explanations of intentional side-effects avoided morality and focused instead on more abstract aspects of the cases. According to one theory, people think costs are generally incurred intentionally. Harming the environment and decreasing sales in New Jersey are both *costs*, because, other things being equal, they would be avoided; it is only for the sake of some overall good (here, increased profits) that they are endured. Helping the environment, on the other hand, is no cost to be endured whatsoever! This asymmetry in accepting costs is purported to explain the asymmetry in intentionality judgements.

However, this *trade-off hypothesis* fails to accommodate other side-effect cases. Consider the case of a terrorist leader who wants to bomb

a nightclub in order to kill Americans. His henchman informs him that it will kill a lot of Australians too. The leader responds, "While it will be good to harm Australians, I really just care about killing Americans. Let's bomb the nightclub!" Here most people conclude that the terrorist intentionally killed Australians. But this is no cost from the terrorist's perspective, since, other things being equal, he'd prefer to kill some Australians. Nor is it a cost from our perspective, since for us killing Americans is no overall good for which it's worth enduring the deaths of Australians. So, an account that equates intentional side-effects with costs fails to explain all cases.

According to another theory, people typically identify a core character or deep self – an aspect of a person's psychology constituted by his fundamental and enduring values and attitudes. When the effects of one's actions are concordant with this deep self, those effects will be judged to have been intentionally brought about. Both Jules and Jill express utter disregard for the environment, so, perhaps they have enduring negative attitudes towards nature. The effect of harming the environment (but not helping it) is consonant with such an attitude, so Jules is thought to have intentionally harmed the environment (but Jill is not judged to have intentionally helped it) according to this *deep self model*. The model also accommodates intentionality judgements in the terrorist case, since the terrorist deeply hates all freedom-loving peoples, including Australians.

Nonetheless, it is tough to see why decreasing sales in New Jersey should be thought consonant with Hal's deep self, or even why one should be thought to hold enduring

attitudes about such things at all. That is, it is tough to see how the deep self model can explain why we think Hal intentionally decreased sales in New Jersey. And the view seems even less likely to explain our intentionality judgements regarding a foreman of a sprocket company, whose chief engineer just explained to him that the new manufacturing process will make the sprockets stronger, but will also make them black (whereas most companies produce sprockets blue or lighter in colour). This foreman says, “I don’t care at all about the colour of our sprockets. It’s important only that we make strong sprockets. Go ahead with the plan.”

People think it more appropriate to say such a foreman intentionally made his sprockets darker than blue when that colour departs from the industry norm. If making the sprockets black actually conforms to the industry norm, few people think the foreman made them black on purpose. But surely the belief that sprockets should be darker than blue is not a fundamental value for one foreman as opposed to another. Thus, like previous accounts, the deep self model succumbs to further vignette testing.

We could go on. There are many more theories we

might discuss. According to some, instances of departure from the norm (either statistical or moral) are what occasion judgements of intentionality in side-effect cases. Another theorist claims that we think a side-effect is intentionally performed in so far as we suppose that the agent would have acted differently if she had fully considered the implications of the side-effect. These theories each have their attendant counter-examples. Or, if they do not yet, experience suggests that they soon will. But rather than discussing more explanations in any detail, it is perhaps more helpful for us to pause and reflect on where this all began and on why it has led us here.

We began by trying to answer a seemingly simple question: What makes us think Jules harmed the environment on purpose? But none of the explanations we’ve considered generalised to the burgeoning class of intentional side-effect cases. But why should an account of a particular kind of intentional side-effect be rejected because it fails to apply to another kind of intentional side-effect?

As we have seen, experimental philosophers have generally attempted to account for all kinds of intentional side-effect cases through elegant theories that appeal to only one or two factors supposedly common to all side-effect cases. >>>



But why should we expect the pattern of intentional side-effect cases to admit of a straightforward explanation, appealing to only one or two factors? Why should we suppose some uniform feature runs through all of the side-effects deemed to have been brought off intentionally? Perhaps it's time to give up on the ideal of a unified explanation.

Perhaps it's time to give up on the ideal of a unified explanation

The desire for parsimony – to posit as few explanatory features as possible – has a rich philosophical history and is often given lots of weight in philosophical theory construction. But, as the psychologist Tania Lombrozo has demonstrated, our bias in favour of parsimony can lead us to adopt simple explanations even when it's far more likely that a complicated explanation is correct.

Lombrozo had participants diagnose the condition of an extra-terrestrial patient, Treda, who had sore minttels and purple spots. Treda might be suffering from one or more of three conditions: 1) Morad's disease, which causes sore minttels; 2) a Humel infection, which causes purple spots; 3) or Trichet's syndrome, which causes both sore minttels and purple spots. She then manipulated the degree to which the three diseases were present in the alien population.

Even when it was twice as likely (given disease distributions across the population) that Treda suffered from both Morad's disease *and* a

Humel infection, two-thirds of people thought Trichet's syndrome was a better explanation for Treda's symptoms. When it was ten times as likely that the more complicated explanation was correct, a third of people still favoured the simple explanation. Perhaps the present status of the intentional side-effects controversy is due to our continued wish for simplicity in the face of mounting evidence that a more complicated story is needed.

The emphasis on parsimony is no doubt driven by old philosophical tendencies. And perhaps one could argue that parsimony should be given significant weight when we are investigating *metaphysical* questions. But it shouldn't be the driving concern for a theory of *folk psychological* judgements.

Neither natural selection nor cultural evolution requires an intelligent designer with an elegant plan. There are undoubtedly features of each intentional side-effect case that lead us to conclude that a particular side-effect was intended. But we shouldn't suppose those are the same features in each case. To adopt this perspective is not to say that previous explanations of intentional side-effects are without merit. Rather we should see previous theories as elucidating different variables at play in intentionality judgements.

One may ask, if we can't identify some limited set of common characteristics shared by all intentional side-effects, how can we ever really explain intentional side-effects at all? I think this worry is ill-founded. We need to recognise that there are at least two ways of fully explaining any pattern of judgements: We can *either* identify a set of features that

invariably lead people to judge that a side-effect was brought about intentionally *or* we can explicate some general method that will allow us to churn out whatever diverse features lead to the intentionality judgement for any intentional side-effect case with which we are presented.

Theorists have primarily pursued the first path so far, and this has led them to optimistically wish for a simple set of common properties to make their task tractable. I want to suggest that we pursue the second path. And I believe we can make some progress on the task of devising a general method for identifying the causes of intentional action judgements by asking what does a person need a concept of intentional action for anyway?

What use is it to judge that Jules (or anyone else) acted intentionally? I cannot fully answer that question. But I want to suggest that having judged that Jules intentionally harmed the environment, I can infer that Jules is likely to harm the environment again in the future. If we begin to see the ordinary concept of intentional action as serving this role – as standing as a marker for actions we think

individuals are likely to perform again – we begin to see why there are different routes to a concept of intentional action.

There are presumably different subtle cues to what any individual in any specific case is likely to do again in the future. And these may extend far beyond psychological facts about the agent in the story or about us when we are reading it to include more abstract features of the agent's situation.

For example, it may be part of the story of why Jules is likely to harm the environment again – and thus, why we judge that she intentionally harmed the environment this time – that many corporate endeavours *do* harm the environment. Jill is unlikely to help the environment again – and so we should not mark Jill as having intentionally helped the environment this time – in part because so few corporate endeavours actually do help the environment.

Of course, there is much detail left out of this brief suggestion. And, in any case, it is inadvisable to take any theory as the whole story when it comes to explaining intentional action – or so experience suggests.

