Francis, fatalism and the fundamental attribution error: A reply to Philip Darbyshire

John Paley *

8 Farm Place, Henton, Chinnor, Oxon OX39 4AD, United Kingdom

SUMMARY

Philip-Darbyshire has apparently gained the impression that I used 'gee-whiz' social psychology to 'absolve poor or negligent practice from any hint of personal responsibility and accountability'. In this reply, I first discuss the fundamental attribution error, which Darbyshire trivialises as the 'fundamental arrogance error' and a piece of 'linguistic puffery'. Second I examine Darbyshire's rather naive view of causation. Third, I suggest that, on a more defensible view of causality, situations as causes do not absolve nurses from 'any hint' of personal responsibility, but they do set limits to responbility, even if Darbyshire finds this morally incovenient. Finally, I argue that the focus on individual accountability should give way to a public health model of containment and prevention. I conclude with some remarks on 'outsider disbelief'.

© 2014 Elsevier Ltd. All rights reserved.

Preamble

The Francis Report raises some unexpected questions about how we account for institutional behaviour in the health services, and in particular how we explain appalling care in hospitals. In the post-Francis discussion there are echoes of debates in other disciplines, especially social psychology. In explaining other people's behaviour, how much weight do we assign to their dispositions – character, attitudes, values – and how much weight do we assign to the situation in which they found themselves? What can psychologists tell us about the causal effects of situational factors? To what extent is helping behaviour influenced by the environment, and to what extent is it merely a reflection of someone's character? Is it true that we are inclined to overestimate the significance of 'what the person is like' and underestimate the significance of 'what the situation was like'? I think these are important questions. In indicating why I think Philip Darbyshire has got them wrong, I am not engaging in a private argument. I am suggesting that psychology provides us with the resources – and an incentive – to look at the Mid Staffs affair from a challenging and unfamiliar angle.

Introduction

Like Rolfe and Gardner (2014), Philip Darbyshire rips into my editorial (Paley, 2013) without stopping to wonder whether a 1200 word

opinion piece is really going to tell the whole story (Darbyshire, 2014). I won't, in this reply, repeat my comments (Paley, in press) on the apparent eagerness to slash and burn, although I will mention the paper in which a fuller account is provided (Paley, 2014a), together with an essay which sketches in the wider philosophical background (Paley, 2014b). Instead, I will examine the basis for Darbyshire's 'impression' that I have used 'gee whizz' social psychology to 'absolve poor or negligent practice from any hint of personal responsibility and accountability' (p. 888).

The reply has three sections. In the first, I will discuss the fundamental attribution error (FAE), an idea which Darbyshire trivialises as the 'fundamental arrogance error' and a piece of 'linguistic puffery'. It is evident from his remarks that he does not really understand what the FAE is. The idea is that, when accounting for people's behaviour, we tend to invoke character traits, attitudes and values, even when the behaviour is constrained by circumstance. We over-emphasise dispositions, we under-emphasise context. The FAE is a theory of attribution. It says: 'we over-attribute behaviour to traits'. Darbyshire, however, seems to think it is a theory of behaviour, as if it said: 'situations explain behaviour, character doesn't come into it'. Apparently, he does not see the difference between 'situations matter more than we think' and '*only* situations matter'.

This misunderstanding of the FAE is compounded by a rather naïve view of causation. According to Darbyshire, claiming that behaviour is constrained by situations makes me a determinist – a 'fatalist' unable to account for nurses who 'do not all behave and respond identically'. In the second section, I will suggest that this view presupposes a theory

* Corresponding author. Tel.: +44 1786 351905; +44 7712 210778.
  E-mail address: john.paley@btinternet.com.

of causality according to which 'causes' are sufficient conditions. I will explain why this theory doesn't work, and sketch an alternative *difference-making* and *probabilistic* account.

In the third section, I will propose that, on a more defensible view of causality, situations-as-causes do not absolve nurses from 'any hint' of personal responsibility and accountability. However, they do set *limits* to responsibility, even if Darbyshire finds this morally inconvenient. I will further suggest that the probabilistic view of causation implies, not a focus on individual responsibility and blame, but a public health model of containment and prevention.

The Fundamental Attribution Error

Let us first be clear about one thing. The FAE is 'not a contention that situations are all powerful, or even that situational influences on behaviour are more powerful than dispositional influences. Rather, it is the contention that the layperson's intuitions give more weight to dispositions and less weight to situational influences than what psychologists have learned… is warranted' (Gilovich and Eibach, 2001, p. 24). As a result of the controversy provoked by the expression 'fundamental attribution error', Ross (who coined it) subsequently preferred to use the term 'lay dispositionism' (Ross, 2001). A related term is 'correspondence bias' (Gawronski, 2004).

In the rest of this section, I will review the evidence for the FAE, and outline some of the more recent theoretical and empirical developments. I will then consider what Darbyshire says about the FAE, and suggest that he has misunderstood it.

The Evidence

The evidence for the FAE is such that Jones (1990, p. 138) called it 'a candidate for the most robust and repeatable finding in social psychology'. The consistency of the evidence extends across a number of different experimental paradigms, all of which suggest that subjects routinely attribute other people's behaviour to their attitudes (Jones and Harris, 1967), opinions (Gilbert and Jones, 1986), personality (Miller et al., 1981), friendliness (Fleming and Darley, 1993), anxiety (Snyder and Frankel, 1976), or moral values (Bierbrauer, 1979), even when situational constraints are manifestly imposed on the person about whom the inference is made. Surprisingly, the FAE is even observed when the experimental subject herself imposes the situational constraints on that person's behaviour (Gilbert and Jones, 1986). For reviews of the literature, see Ross and Nisbett (2011), Gilbert and Malone (1995), Gawronski (2004), and Reeder (2009).

In one classic study (Jones and Harris, 1967), for example, subjects inferred that students who had written essays favourable to Castro had pro-Castro attitudes, even though they were informed that the student assignment had been to write an essay defending him. Studies reporting similar results include: Gilbert and Jones (1986), Ross et al. (1977), and Kunda and Nisbett (1986). The literature confirms that findings of this sort are independent of the experimental subject's own attitudes (Alicke et al., 1996), warnings of judgmental bias (Croxton and Miller, 1987), further information about the essay writer (Ajzen et al., 1979), different types of information about the constraint (Croxton and Morrow, 1984), and artificial or authentic essays (Miller et al., 1990).

The previous paragraph cites research in the attitude attribution paradigm; but in the present context, the moral attribution paradigm has even greater relevance. In this paradigm, participants first learn about social psychology experiments in which situational factors were clearly implicated in the experimental subjects' behaviour (for example, Milgram, 1963), and are then invited to make inferences about people who are in situations similar to those defined by the experimental protocol (Bierbrauer, 1979). Typically, the participants attribute immoral dispositions to these people – they are bad people – even though they are in situations known to influence behaviour strongly. In the Milgram example, participants infer that Milgram's subjects were unusually cruel, or were hostile types, or had blindly obedient personalities (Miller et al., 1974; Pietromonaco and Nisbett, 1982). If reminded that his sample was large and diverse, they infer that everybody must be like that.

A variant of this paradigm asks participants to predict what experimental subjects will do, or to predict what people in a similar situation will do. Typically, participants overestimate the likelihood that people will behave 'morally', which again suggests that they are failing to allow sufficiently for situational constraints. Milgram (1965) himself invited 40 psychiatrists to predict the performance of subjects in his experiment. Their predictions were wildly inaccurate. 62% of Milgram's subjects went all the way, to the highest level of shock; the psychiatrists predicted that only 0.01% would. As Milgram says, this is a 'whopping discrepancy' (p. 72).

Theoretical and Empirical Developments

The original formulation of the FAE referred to the relative weights that people attach to dispositions and situations. However, the incorporation of the social cognition perspective (Carlston, 2013) has led to 'stage models' of attribution, according to which an initial dispositional inference might be adjusted in order to take account of situational factors (Reeder, 2013). In these 'anchoring and adjustment' models, the initial stage is a spontaneous trait inference, which is unconscious and automatic. It can be followed, though not inevitably, by a situational adjustment stage, which requires conscious cognitive resources (Gilbert, 1989). At this point, as I observed in Paley (in press), the situationist tradition begins to overlap with the literature on automaticity and the cognitive unconscious (Kihlstrom, 1987; Hassin et al., 2005).

The correction for situational factors does not necessarily follow the dispositional inference. It has been suggested, for example, that people check the diagnostic value of behaviour before making a situational correction, if indeed they get this far (Reeder, 1993, 2001; Reeder and Brewer, 1979). That is, they first evaluate the degree to which the behaviour they observe is diagnostic of a particular character trait, and only then – and only subject to certain conditions – take situational factors into account. Specifically, if the diagnostic value of the behaviour *b* for character trait *t* is low, an initial dispositional inference may be amended on the basis of situational factors. If, on the other hand, the diagnostic value of behaviour *b* for trait *t* is high, a correction based on the situation will not be triggered, and the initial dispositional inference will not be modified.

For example, many of those commenting on Milgram studies regard going to the highest level of shock as immoral (Safer, 1980). Equally, they regard immoral behaviour as highly diagnostic of being a bad person. So the dispositional inference – that Milgram's subjects are cruel, hostile, or blindly obedient – is not corrected for situational considerations. The schema underlying this inference is that only immoral people behave immorally. So the diagnostic value of immoral behaviour for the trait 'bad person', or 'cruel disposition', is high. Hence, the second-stage situational correction does not occur.

The primacy of the diagnostic check is particularly apparent in cases of moral attribution; and the worse the behaviour, the less likely it is that a situational correction will be activated: 'the more troublesome or threatening the behavior that observers attempt to explain, and the more extreme the actions with which they are concerned, the more tempting it is to attribute primary responsibility to disagreeable or damaged "others" whose bad acts are thought to be the products of their flawed characters and ill-advised choices' (Haney and Zimbardo, 2009, p. 807). If this is right, even those who in general accept that situational forces constrain behaviour may still act on the assumption that immoral behaviour is a 'sufficient indicator of an immoral disposition' (Gawronski, 2004, p. 203).

A recent theoretical proposal approaches the FAE from a different, but compatible, angle. This is the suggestion that correspondence bias is a form of the 'inherence heuristic' (Cimpian and Salomon, 2014).