

Correspondence

Psychological Medicine, 46 (2016).

doi:10.1017/S0033291716001793

First published online 15 August 2016

Causes and correlates of intrusive memory: a response to Clark, MacKay, Holmes and Bourne

No reader of *Psychological Medicine* needs to be reminded that correlation does not imply causation, but there are two attitudes that might be taken to that fact, and between them a spectrum of positions. I suspect that those holding positions towards either end of this spectrum overestimate their view's preponderance, and that this explains some of the mutual misunderstandings in our debates concerning psychological disorder. An example may be found in a recent exchange between me, on one hand, and Clark, MacKay, Holmes and Bourne, on the other (Clark *et al.* 2016b; Mole, 2016). Before considering that exchange, we should consider the two contrasting attitudes to correlation and causation. Most feasible positions lie somewhere between these extremes, but opinions differ as to whether the most sensible lie nearer the first or second.

Our first attitude takes the maxim about correlation and causation as warning of a fallacy to which humans are prone. Just as we need to avoid the cognitive illusions that lead us to ignore the base rate when assessing posterior probabilities, or to treat as most likely that which is most typical (Kahneman & Tversky, 1973, 1974), so we must avoid succumbing to our tendency for treating correlations as causal (Matute *et al.* 2011). These illusions look persuasive, even to those in the know (Diaconis & Freedman, 1981), but they are nonetheless fallacious. For causal inferences to be better than illusory, such inferences require not only that variables are correlated: They also require some theory of the mechanism by which the alleged causation is implemented. If the study of psychological medicine is to take a properly scientific approach to its causal hypotheses then it must, on this view, concern itself with *mechanisms*.

Many theorists are sympathetic to at least some part of this attitude (Revonsuo, 1999; Poland & Von Eckardt, 2004; Kendler, 2008; Craver, 2009). Among readers of this journal it might even be orthodox (Kendler *et al.* 2011). But elsewhere one finds a different attitude being taken; which also can be approbated plausibly.

This second attitude sees correlation's failure to imply causation as a symptom of the fact that scientific inference is not deductive (Popper, 1959). Observations of correlation do not entail causal claims, but nor (for example) did Eddington's eclipse photographs entail the general theory of relativity. Scientific data never support theories by entailing them (Poincaré, 1902). Instead a theory gets supported by the data by giving a parsimonious explanation for them (Lipton, 1991). The fact of some variables being correlated calls for an explanation: the theory according to which there is a causal relation between those variables frequently gives the most parsimonious explanation available (Reichenbach, 1949). Correlation, on this view, does not imply causation, but does provide appropriate evidence for it, all else being equal. That evidence may not answer the Humean skeptic, but he does not set the standard by which science should be judged.

These two attitudes recommend opposite reactions when a move is made from correlational data to a causal claim. The first sees this as a fallacy, the second as a legitimate inference. These attitudes are, therefore, in tension – and the position of a theorist who attempted to adopt both would be precarious – but we should avoid choosing between them. Both contain some truth. The first is right that we should be wary of our susceptibility to cognitive illusions; the second that we should not hold scientific inference to the standard of deductive proof. Sensible views must acknowledge both points, but within the scope of the sensible there is room for significant disagreements of emphasis. My own sympathies tend towards the first attitude. Clark *et al.*'s (2016b) response to my letter suggests that theirs tend to the second. Lacking a theory of the mechanisms by which brain states and mental states interact, I take correlations to give us only a tentative picture of the causal structures in which those states participate. Clark *et al.* (2016a, b) treat that picture as if it were complete.

The resulting disagreement comes into focus when it is seen against the background of our more fundamental agreement. Clark *et al.* (2016b) repeatedly note the centrality of the point over which we agree, remarking that my critique of their methodology:

has no bearing on [their] main analyses nor the interpretations and conclusions that are drawn – namely that there is a specific peri-traumatic pattern of brain activation that predicts intrusive memory formation. (p. 1787, emphasis added, here and in the two following quotations)

They summarize this position by saying that:

the main aim of the original paper (and our subsequent replication) was to investigate whether the occurrence of an intrusive memory is *determined by* the neural activity during the original encoding of experimental trauma. (p. 1788)

And the conclusion of that replication, as given in its abstract, is that:

The left inferior frontal gyrus may be *implicated in* both the encoding and involuntary recall of intrusive memories (Clark *et al.* 2016a, p. 505).

I unreservedly agree with these authors' claim that the involuntary recall of intrusive memories is not to be explained by reference only to processes taking place in the period *after* an intrusively recollected event has been experienced: the explanation of intrusiveness must partly be sought in the events that take place when an intrusively remembered episode is first encoded. On this point the authors' evidence is flawless. Our dispute concerns what it is, in the initial experience of an episode that creates its propensity for involuntary recollection. Clark *et al.* (2016a, b) hold that the causes of this propensity are not wholly emotional. I maintain that their reasons for holding this are flawed.

More specifically, I think those reasons require too liberal an attitude to the move from correlation to causation. That attitude can already be seen in the passages quoted above. The first makes a point about *prediction*, and so can be made true by a mere correlation. The third makes a claim about a brain area being *implicated in* the relevant psychological phenomena, and so requires this area's involvement in a causal operation. The unmarked move between these is legitimate only if our second attitude to causal inference is regnant. These authors also remark that 'Bourne *et al.* (2013) did not claim that the secondary analysis was a mathematical proof, nor that it meant that emotion played no role in intrusive memory encoding'. In taking this disavowal of mathematical proof to be relevant, they again seem to be adopting the second attitude.

Before asking whether their position requires an implausibly strong version of that attitude, an additional point should be noted (since the last of the above quotations raises it), which concerns the issue that is supposedly in dispute. In that last quotation we see the authors insisting that their work 'did not claim [...] that emotion played no role in intrusive memory encoding'. They also write:

Mole (2016) argues that there are distinct emotion-based and cognitive-based hypotheses for the formation of intrusive memories. He states: 'Bourne *et al.* suggest that the evidence given in their 2013 study favours hypotheses of this second type' – referring to a favouring of cognitive processing over emotional processing for the formation of intrusive memories. (Clark *et al.* 2016b, p. 1788)

They 'strongly disagree' with the claim that they take me to attribute to them, saying that:

Unlike Mole (2016), we do not argue that these are distinct hypotheses and explanations. Instead, we suggest the evidence taken together strongly implies that emotional processing is important for intrusive memory formation but that other factors in addition to emotion are also important (as we have argued elsewhere, e.g. Holmes & Bourne, 2008).

These remarks embody a misreading. The claim with which they disagree is not among the claims I made, nor among those that I attributed to anybody else. Nobody denies that emotion is important to the explanation of intrusiveness. As these authors say (and as I quoted them as saying) their position is that 'although intense emotional reaction may be a necessary condition for flashback formation it appears not to be sufficient'. My point was that their evidence fails to support this claim of insufficiency, not that it fails to support some other claim, according to which emotion plays no role. None of the hypotheses that I considered 'favoured cognitive processing over emotional processing for the formation of intrusive memories'. Those hypotheses were (1) that: 'intrusively remembered episodes differ from other episodes *just because* [of] the emotional responses that they elicit, at the time when they are first experienced' (Mole, 2016, p. 1785, emphasis added); and (2) that 'the initial processing of intrusively remembered episodes [is] marked, *not only by emotion, but also by some more purely cognitive peculiarity*' (Mole, 2016, emphasis added). The first hypotheses cite emotion alone as the cause of intrusiveness; the second cite emotion *plus also* some non-emotional factor. Neither says that emotion 'plays no role'. Neither 'favours cognitive processing over emotional processing'. The first say that every process that contributes to the difference between intrusive and normal memory is an emotion-constituting process; the second that *some* of these processes are non-emotional. Clark *et al.* (2016b) take the second position:

We wrote that 'although intense emotional reaction may be a necessary condition for flashback formation it appears not to be sufficient' in Bourne *et al.* (2013; p 1529), by which we intended to suggest that intrusive memories [*sic*] are not formed solely due to extreme emotion, but also due to a number of other factors. In other words, heightened [*sic*] emotion is necessary, but alone it is not sufficient, requiring the involvement of other cognitive processes for intrusive memory formation.

This position is plausible, but my contention continues to be that the studies in question provide no justification for believing it. Those studies involve an analysis of functional magnetic resonance imaging data that is said to provide 'two key pieces of information'. We

can consider the evidential status of both, taking the second piece first.

This second piece of information is that ‘the left inferior frontal gyrus and middle temporal gyrus show activation that is not compatible with simple increasing levels of emotionality – activity in these areas is lower during Potential scenes than during both Control and Actual (Intrusive) scenes’ (Clark *et al.* 2016*b*, p. 1788). To support Clark *et al.*’s (2016*a, b*) position this information would need to weigh against the hypothesis according to which every process contributing to the intrusiveness of memories is an emotion-related process. That cannot be done by showing that some of the centres that *are* involved in emotion work less hard when processing intrusive scenes than when processing scenes that will be recollected normally.

To see this, consider an analogy. Suppose that diners in some restaurant occasionally become unwell, and that a consensus has been reached that the causes of this are to be found in processes that occur before their food arrives at the table. One hypothesis is that the processes responsible are those happening in the kitchen. An alternative is that things happening in the kitchen are insufficient, and that the front of house staff must also play a role. The hypothesis implicating the front of house staff would gain no support from observing that *some* kitchen staff do the same things, whether the diner becomes ill or not, nor from observing that some kitchen staff work less hard on occasions when the diner does become ill. Similarly, when faced with the fact that some memories become intrusive (and with a consensus that events at the time of encoding are implicated in this) there is one hypothesis saying that the processes responsible are in every case emotional, and another saying that non-emotional processes must also play a role. The latter hypothesis gains no support from observing that certain emotional processing areas do the same things, whether the memory becomes intrusive or not, nor from observing that some such areas are less active on occasions when the memory does become intrusive. If we had evidence that *everyone* in the kitchen does *the same thing* when diners get ill and when they do not, then the situation would be different, but to suppose that we have evidence of that sort in the psychological case is to suppose that the causal picture given by the correlational data is complete. This requires that we endorse the second attitude in its strongest form.

Clark *et al.*’s (2016*b*) other ‘key piece of information’ is said to be that ‘many brain regions not associated with emotional processing are implicated (with the acknowledged caveat of the limitations of reverse inference)’ (p. 1787). Their entitlement to that ‘implicated’ is the point in contention. Their evidence for it comes from an analysis showing that the activity of

regions not associated with emotion carries information on the basis of which predictions about intrusiveness can be made. All sorts of causal arrangements might account for the accuracy of these predictions. The activity of my barometer carries information on the basis of which predictions can be made about the occurrence of a storm. This does not show the barometer to be implicated in the storm’s production. At most it shows that there is some causal nexus in which barometer and storm both participate.

The finding that some brain region responds differentially to intrusive and non-intrusive scenes would imply that something other than emotion contributes to intrusive recollection only if the differentially responding region played some role in bringing about intrusive recollection, and only if the region in question played no role in emotion. Those who favour the first attitude to causal inference will take neither claim to be well supported by correlational data.

Acknowledgements

Thanks to Belen Martin-Barragan, and especially to Emily Holmes, for help, advice, and information. This work was supported by an Insight Grant from the Social Sciences and Humanities Research Council of Canada (grant number 12R22982).

Declaration of Interest

None.

References

- Bourne C, MacKay CE, Holmes EA (2013). The neural basis of flashback formation: the impact of viewing trauma. *Psychological Medicine* **43**, 1521–1532.
- Clark IA, Holmes EA, Woolrich MW, MacKay CE (2016*a*). Intrusive memories to traumatic footage: the neural basis of the encoding and involuntary recall. *Psychological Medicine* **46**, 505–518.
- Clark IA, MacKay CE, Holmes EA, Bourne C (2016*b*). A reply – acknowledged reasonable limitations in a secondary analysis but key conclusions remain in ‘The neural basis of flashback formation: the impact of viewing trauma’. *Psychological Medicine* **46**, 1787–1789.
- Craver CF (2009). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press: Oxford.
- Diaconis P, Freedman D (1981). The persistence of cognitive illusions. *Behavioral and Brain Sciences* **4**, 333–334.
- Holmes EA, Bourne C (2008). Inducing and modulating intrusive emotional memories: a review of the trauma film paradigm. *Acta Psychologica* **127**, 553–566.
- Kahneman D, Tversky A (1973). On the psychology of prediction. *Psychological Review* **80**, 237–251.
- Kahneman D, Tversky A (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**, 1124–1131.

- Kendler KS** (2008). Explanatory models for psychiatric illness. *American Journal of Psychiatry* **165**, 695–702.
- Kendler KS, Zachar P, Craver CF** (2011). What kinds of things are psychiatric disorders? *Psychological Medicine* **41**, 1143–1150.
- Lipton P** (1991). *Inference to the Best Explanation*. Routledge: London.
- Matute H, Yarritu I, Vadillo MA** (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology* **102**, 392–405.
- Mole C** (2016). A methodological flaw in ‘The neural basis of flashback formation: the impact of viewing trauma’. *Psychological Medicine* **46**, 1785–1786.
- Poincaré H** (1902). *La Science et L’Hypothèse*. Flammarion: Paris.
- Poland JS, Von Eckardt B** (2004). Mechanism and explanation in cognitive neuroscience. *Philosophy of Science* **71**, 972–984.
- Popper K** (1959). *The Logic of Scientific Discovery*. Hutchinson: London.
- Reichenbach H** (1949). *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. The University of California Press: Los Angeles, CA.
- Revonsuo A** (1999). Neuroscience and the explanation of psychological phenomena. *Behavioral and Brain Sciences* **22**, 847–849.

C. MOLE

Department of Philosophy and Cognitive Systems, University of British Columbia, 1866 Main Mall, E370, Vancouver, BC, V6T 1Z1, Canada

Author for correspondence: Christopher Mole, Department of Philosophy and Cognitive Systems, University of British Columbia, 1866 Main Mall, E370, Vancouver, BC, V6T 1Z1, Canada.

(Email: cmole@mail.ubc.ca)