

# Not so easy to spot – a failure to replicate the Macbeth effect across three continents

'Out, damned spot!' cries a guilt-ridden Lady Macbeth as she desperately washes her hands in the vain pursuit of a clear conscience. Consistent with Shakespeare's celebrated reputation as an astute observer of the human psyche, a wealth of

contemporary research findings have demonstrated the reality of this close link between our sense of moral purity and physical cleanliness.

One manifestation of this was nicknamed the Macbeth effect - first documented by Chen-Bo Zhong and Katie Liljenquist in an influential paper in the high-impact journal *Science* in 2006 - in which feelings of moral disgust were found to provoke a desire for physical cleansing. For instance, in their second study, Zhong and Liljenquist found that US participants who hand-copied a story about an unethical deed were subsequently more likely to rate cleansing products as highly desirable.

There have been many 'conceptual replications' of the Macbeth effect. A conceptual replication is when a different research methodology supports the proposed theoretical mechanism underlying the original effect. For example, last year, Mario

Gollwitzer and André Melzer found that novice video gamers showed a strong preference for hygiene products after playing a violent game.

Given the strong theoretical foundations of the Macbeth effect, combined with several conceptual replications, University of Oxford psychologist Brian Earp and his colleagues were surprised when a pilot study of theirs failed to replicate Zhong and Liljenquist's second study. This pilot study had been intended as the start of a new project looking to further develop our understanding of the Macbeth effect. Rather than filing away this negative result, Earp and his colleagues were inspired to examine the robustness of the Macbeth effect with a series of direct replications. Unlike conceptual replications, direct replications seek to mimic the methods of an original study as closely as possible.

Following best practice guidelines, Earp's team contacted Zhong and Liljenquist, who kindly shared their original materials.

Another feature of a high-quality replication is to ensure you have enough statistical power to replicate the original effect. In psychology, this usually means recruiting an adequate number of participants. Accordingly, Earp's team recruited 153 undergraduate participants - more than five times as many as took part in Zhong and Liljenquist's second study.

Exactly as in the original research, the British students hand-copied a story about an unethical deed (an office worker shreds a vital document needed by a colleague) or about an ethical deed (the office worker finds and saves the document for their colleague). They then rated the desirability and value of several consumer products. These were the exact same products used in the original study - including soap, toothpaste, batteries and fruit juice - except that a few brand names were changed to suit the UK as opposed to US context. Students who copied the unethical story rated the desirability and value of the various hygiene and other products just the same as the students who copied the ethical story. In other words, there was no Macbeth effect.

It's possible that the Macbeth effect is a culturally specific phenomenon. Next, Earp and his team conducted a replication attempt with 156 US participants using Amazon's Mechanical Turk survey website. The materials and methods were almost identical to the original except that participants were required to re-type and add punctuation to either the ethical or unethical version of the office worker story. Again, exposure to the unethical story made no difference to the participants' ratings of the value or desirability of the consumer products - with just one anomaly: participants in the unethical condition placed a higher value on toothpaste. In the context of their other findings, Earp's team think this is likely to be a spurious result.

Finally, the exact same procedures were followed with an Indian sample - another culture that like the US, places high value on moral purity. Nearly three hundred Indian participants were recruited via Amazon's Mechanical Turk, but again no effect of exposure to an ethical or unethical story was found on ratings of hygiene or other products.

Earp and his colleagues want to be clear - they're not saying that there is no link between physical and moral purity, nor are they dismissing the existence of a Macbeth effect. But they do believe their three direct, cross-cultural replication failures call for a 'careful reassessment of the evidence for a real-life "Macbeth Effect" within the realm of moral psychology'.

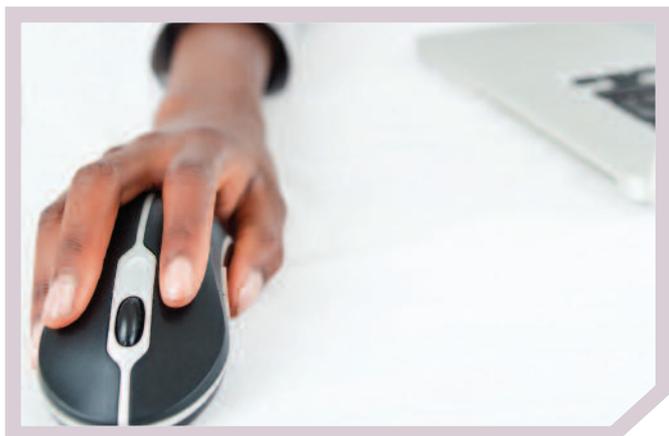
This study, due for publication next year, comes at time when reformers in psychology are calling for more value to be placed on replication attempts and negative results. 'By resisting the temptation ... to bury our own non-significant findings with respect to the Macbeth Effect, we hope to have contributed a small part to the ongoing scientific process,' Earp and his colleagues concluded.

## Further reading

An unsuccessful conceptual replication of the Macbeth effect was published in 2009 ([www.jasnh.com/pdf/Vol6-No2.pdf](http://www.jasnh.com/pdf/Vol6-No2.pdf)). Later, in 2011, another paper failed to replicate all four of Zhong and Liljenquist's studies, although the replications may have been underpowered ([www.ncbi.nlm.nih.gov/pubmed/21568173](http://www.ncbi.nlm.nih.gov/pubmed/21568173)).



In *Basic and Applied Social Psychology*



## Questioning the effectiveness of headlines?

In *Social Influence*

In the competition for readers' mouse clicks, a favoured trick is to phrase headlines as questions. As a way to grab attention, question headlines have been recommended by editors for decades. But what is new, is the ability to measure how often readers choose to click a headline. For a new paper, researchers in Norway have used Twitter to find out whether question headlines really do entice more clicks.

Linda Lai and Audun Farbroth used a science communication Twitter feed that had 6350 followers. Real stories were tweeted to these followers twice, an hour apart. The first tweet used a statement headline, such as 'Power corrupts'. The second, referring to the same story, was a question, either self-referencing ('Is your boss intoxicated by power?') or non-self-referencing ('Are bosses intoxicated by power?').

Lai and Farbroth found that self-referencing question headlines were clicked on

average 175 per cent more often than statement headlines (this advantage dropped to 150 per cent for non-self-referencing question headlines). The difference in clicks for question and statement headlines was statistically significant, but the difference between the self-referencing and non-self-referencing headlines was not.

A follow-up study conducted via an online auction site yielded similar results, though this time the difference between the two types of question headline was statistically significant.

A potential criticism is that some of the headlines may have differed in other ways besides their quizzical status – such as in the example about bosses, where question headlines referred to 'intoxication' but the statement headline did not. Assuming questions really do provoke more clicks than statements do, another weakness of this paper is that it doesn't tell us anything about why this is the case.



## Language and the brain – which side are you on?

In the *Journal of Neurolinguistics*

Simple facts about the brain are rare, but one of them is that for most people language function is located mainly in their left brain hemisphere. The stats vary according to the measures used, but this is the situation for around 95 per cent of right-handers and approximately 75 per cent of left-handers. When it comes to the brain though, few things are straightforward.

If we dig deeper, as Byron Bernal and Alfredo Ardila have done for a new review paper, we find a more complex, two-sided story. A dramatic demonstration of this comes from the Wada test, named after Japanese neurologist Juhn Atsushi Wada. With the patient awake, anaesthetic is injected into the neck or head on one side to effectively shut down function in that side of the brain. Speech and language comprehension tests are conducted first with one hemisphere silenced, then the other. Looking at the results from 1799 Wada tests, most of which were conducted with epilepsy patients prior to surgery, Bernal and Ardila found that 10 per cent of right-handers and 27 per cent of left-handers (and the ambidextrous) showed evidence that their language function was supported by both brain hemispheres.

The way that bilateral language function manifests in the Wada test varies from patient to patient. In some, shutting down one hemisphere has no effect on their language abilities, while shutting down the other only partially interferes with language.

In other patients, shutting down one hemisphere completely impairs language, while shutting down the other also has a partial adverse effect. And in a final group, shutting down either hemisphere results in only a partial impairment to language.

The reason for these different patterns, Bernal and Ardila explain, is that there are various ways that language function can be shared between the hemispheres. Using brain scans from real-life case studies, they show how in some people all functions of language are shared between the left and right brain, whereas for other people some subfunctions of language are bilateral, but not others. Related to this, some people show evidence that the different steps of language function are distributed sequentially between the hemispheres, so there is no redundancy; whereas other people show a kind of parallel arrangement.

Caution is needed when extrapolating from patient studies to healthy people, as it's possible that the brain has altered its function to adapt to disease. This aside, Bernal and Ardila's fascinating review is a reminder of the brain's complexity. The factoid that in most people language is left-lateralised conceals a messy reality. As Bernal and Ardila write: '[L]anguage dominance is mostly a matter of hemispheric advantage for a specific multi-modular cognitive function: language. As such, language in a strict sense is up to a certain point a bilateral brain function.'



The material in this section is taken from the Society's **Research Digest** blog at [www.researchdigest.org.uk/blog](http://www.researchdigest.org.uk/blog), and is written by its editor **Dr Christian Jarrett**. Visit the blog for full coverage including references and links, additional current reports, an archive, comment and more.

Subscribe by RSS or e-mail at [www.researchdigest.org.uk/blog](http://www.researchdigest.org.uk/blog)

Become a fan at [www.facebook.com/researchdigest](http://www.facebook.com/researchdigest)



Follow the Digest editor at [www.twitter.com/researchdigest](http://www.twitter.com/researchdigest)