When Too Much Evidence Can Be Bad

January 15, 2016 | by Stephen Luntz

When everyone agrees on something it can be a sign they're actually wrong. As strange as this may seem, it's something we're familiar with in certain circumstances. Now a scientific paper has explained the concept mathematically, and given a start to the process of working out when to treat unanimity of opinion or evidence with care.

A paper to appear in Proceedings of the Royal Society A (preprint on arXiv) presents three examples of situations where our confidence in an answer should fall as more evidence pointing in the same direction is added.

At first, the idea seems counter-intuitive, but author Professor Derek Abbott of the University of Adelaide points to well-known examples. "When the leader of a country is re-elected with a 100 percent vote we know something is wrong," Abbott told IFLScience. "With a huge population there will always be divergence, so 100 percent is very suspicious."

The paper presents a less obvious example, arguing, "Increasing confirmatory identifications in a police line-up or identity parade can, under certain conditions, reduce our confidence that a perpetrator has been correctly identified."

We might expect that if a dozen people all picked a suspect out of a line-up that this means the police have found the criminal. Indeed, Abbott notes, in some cases this would be right. "The situation would be quite different if the witnesses had all been taken hostage for a month by the suspect," Abbott said in a statement. "Then you would expect them all to agree very well who the kidnapper was."

However, most line-ups involve people trying to identify someone they have seen briefly in the course of a fast-moving incident. Numerous studies have shown that there is a very high error rate in these cases, as high as 48 percent, Abbott said. In such circumstances, if too many witnesses all point to the same suspect, it is more likely to indicate the process was biased, consciously or not. One such biased parade saw white men given dark face make-up and put in a line-up with a black suspect.
The idea has a long history, with the paper pointing to the ancient Israeli Sanhedrin court system, who were forbidden to put a suspect to death when all 23 members agreed on his or her guilt. More recently numerous examples have emerged of why this was a good idea.

The paper provides a mathematical basis to test hunches that unanimous agreement is too perfect, exploring examples where confidence first rises, then falls, as further tests produce the same result.

Abbott told IFLScience that in cases where the likelihood of random errors can be measured, such as in the line-ups example, the work could be used to warn people when outcomes are excessively perfect. Even when such measurements are not possible, he said the idea should prompt us to rethink the quest for consensus among larger groups. “We might want to record dissent rather than trying to bring people around,” Abbott said. “Disagreement is evidence of not having systemic bias.”