In Defense of Fishing

Richard W Byrne
University of St Andrews

Using an example from animal cognition, I argue that the problems of bias—inherent in choosing null hypotheses or setting Bayesian priors—can sometimes be avoided altogether by collecting more and better *observational* data before setting up tests of any sort.

Petrinovich shows us that it is all too easy to 'support' the null hypothesis by carrying out experiments without sufficient power. Combined with the—by now, well known—liability of weakly-supported experimental hypotheses, no-one can doubt the magnitude of the problem with the classical method of statistical analysis. One aspect of it centers on the very idea of a *null hypothesis*. In our statistics classes, we're taught that the null hypothesis should be chosen to reflect the status quo: the currently accepted wisdom which our novel hypothesis might overthrow. But what is accepted wisdom to some scientists may be heresy to others.

Take the case of imitation by apes. Nobody doubts that these animals learn socially to do things that they would not otherwise have done: the controversy has been about how they do it. The behavioral mechanisms of *stimulus enhancement* (Spence, 1937), i.e. seeing another's actions draws attention to a restricted and often key range of stimuli, and *social facilitation* (Zajonc, 1965), i.e. seeing another's actions increases one's own, are insufficient to explain the apparent copying by apes. Two more-powerful mechanisms have been proposed: *impersonation*, sometimes called true imitation (Tomasello, 1996), in which the observer copies the movements it sees used to achieve an effect also desires itself; and *emulation* (Wood, Bruner, & Ross, 1976), in which the observer works towards achieving a goal it has observed another reaching, but uses actions from its existing repertoire to do so. Since we're talking about copying by conspecifics, actions used spontaneously are often very similar to those of others, so the results of these two, logically distinct processes can look more-or-less identical (Tomasello, Davis-Dasilva, Camak, & Bard, 1987). So, which should be the null hypothesis, imitation or emulation? To many animal behavior researchers, who have observed apes in naturalistic enclosures or in the wild, it has seemed obvious that they imitate each other quite often, and learn important skills from doing so: the idea of that this is all mere emulation would have to be proved. In contrast, for many developmental psychologists, imitation *sensu* impersonation is part of a suite of uniquely human abilities that would be remarkable and highly

unlikely in a non-human (Tomasello, 1998). The risk is that whichever we take to be the null will inevitably be 'favored' by the lack of any decent evidence against it (Byrne, 2002).

Taking a Bayesian approach helps a bit but may not get us out of this sort of bind. It would certainly help to force attention on those predictions of the competing hypotheses that genuinely differ. And it would avoid the bias towards retaining the hypothesis we think most acceptable—or would it? Someone still has to set the prior probabilities of each hypothesis in a Bayesian analysis. In the case of ape copying, the two camps *might* just agree on making the priors equal; in other cases, that is highly unlikely. Think of the vexed question of the origin of covid19, and how we would compare these hypotheses: (1) escape from a virus lab, either a result of casual genetic meddling or biological weapon design; (2) species-hopping from a zoological reservoir, facilitated by the horrible conditions of animals held in the repugnant 'wet markets' of many Chinese cities; (3) part of establishing universal mind-control by an evil genius, leading to mass implanting of silicon-chip devices by vaccination. Proponents of all three accounts do exist, and for the first two they include serious scientists; but getting definite evidence against any of them seems quite tricky, so choice of priors is likely to influence the result strongly.

I would like to defend a somewhat heretical suggestion. Journal editors are nearly unanimous in insisting on setting out clearly defined hypothesis in one's introduction so that reviewers can evaluate whether the design and conclusions answer the questions raised. Is this always a good idea? (Quite apart from whether it is always done honestly in advance.) In the case of the ape imitation controversy, the dénouement was in fact from a purely observational study: a 'fishing expedition', in the derogatory phrase. Orangutans at a 'rehabilitation' facility, where carers work to free ex-captive apes from any human attachments or acquired habits, some individuals were noticed—and carefully recorded—impersonating numerous human activities (Russon & Galdikas, 1993), apparently just for the fun of it. These activities included washing clothes, sawing wood (with a stick), tidying leaves on a forest path into neat little heaps, re-kindling a fire with the aid of paraffin decanted into a jug for carrying, canoeing (with a stolen boat), and so forth. These were not activities reinforced by humans in any way, whether in prior captivity or at the release facility: exactly the opposite in the latter case. Nor were they natural actions of orangutans: obviously. The sheer quality and quantity of the observational data was sufficient to rule out any explanation except a natural ability to impersonate.

I wonder, therefore, how often the poor results of weak design and inappropriate use of statistics might have been circumvented altogether by devoting more preliminary effort to observing and understanding the problem before any hypotheses were erected to test. Time was, of course,

when ideas like these would (at least in animal behavior) have seemed mainstream (cf. Tinbergen, 1959 or Van Lawick-Goodall, 1968). But notice the length of those examples, 79 and 131 pages respectively; who nowadays would struggle to publish such things, when they could instead be getting out lots of "REF-worthy[i]", high-impact, Tweetable papers that test some hypothesis or another? I mean, these are serious times!

References

Byrne, R. W. (2002). Emulation in apes: verdict 'not proven'. *Developmental Science, 5*, 20-22.

Russon, A. E., & Galdikas, B. M. F. (1993). Imitation in free-ranging rehabilitant orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 107*, 147-161.

Spence, K. W. (1937). Experimental studies of learning and higher mental processes in infra-human primates. *Psychological Bulletin, 34*, 806-850.

Tinbergen, N. (1959). Comparative studies of the behaviour of gulls (*Laridae*): A progress report. *Behaviour*, *15*, 1-70.

Tomasello, M. (1996). Do apes ape? In C. M. Heyes & B. G. Galef (Eds.), *Social learning in animals: The roots of culture* (pp. 319-346). Academic.

Tomasello, M. (1998). Uniquely primate, uniquely human. *Developmental Science, 1*, 1-30.

Tomasello, M., Davis-Dasilva, M., Camak, L., & Bard, K. (1987). Observational learning of tool-use by young chimpanzees. *Human Evolution, 2*, 175-183.

Van Lawick-Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe Stream Reserve. *Animal Behaviour Monographs*, *1* (Part 3), 161-311

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269-274.

---

[i] For more than 40 years, British academic research has been judged by some sort of recurring process in which staff, their outputs and apparent impact have been judged by a series of committees: the "research assessment exercise" (RAE) and "research excellence framework" (REF) are two incarnations. Since funding follows the judgements, there has been continual pressure to produce outputs that can easily be judged, e.g. by journal impact factor, citations or press coverage.