Re-reading the special issue of *American Statistician* on “Moving to a World Beyond *p* < .05”, I founding myself repeating one of the mantras of behavioral parent training–“focus on what you *want*, not what you *don’t* want.” To be effective, parents are instructed to give children clear positive alternatives instead of just telling them to *stop* misbehaving. This applies as well to clinicians coaching parents–instructing a parent to stop yelling won’t get you as far as showing them a clear alternative–like positively reinforcing a child’s’ good behavior.

Scientists and students seeking more rigorous research results, perhaps, are not so different from parents trying to coax better behavior out of their children. As I re-read the editorial by Wasserstein, Schirm, and Lazar, I nodded along with the authors’ (now-familiar) criticisms of p-values and significance thresholds, while silently asking “what do we do instead?” The authors clearly anticipate such concerns–the first section of their article is called “‘Don’t Is Not Enough”--yet their overview of what to *do* ends up being exponentially longer than their short list of what *not* to do, with much of the advice proving vague or contradictory (as many of the individual authors of the special issue articles do not themselves agree on what should count as best practice in statistics). It wouldn’t surprise me if many scientists have responded to these recommendations roughly the same way as a parent who, shortly after being told to stop yelling at their child, is given a 500-page scholarly book on parenting.

The problem is not the advice itself–which I have no doubt is sound and rigorous–but, to deploy another parent-training mantra: “children [and scientists] respond to incentives.” Just as a child will continue misbehaving in spite of parental commands if their misbehavior is otherwise rewarded, scientists can know internally that a particular practice is flawed or problematic, but will continue engaging in it if they receive professional rewards or benefits from doing so. Or, as Steven Goodman writes in his contribution to the special issues, the “basic explanation” for the universal use of *p*-values is the fact that they are universally used: “It is the same reason we can use money. When everyone believes in something’s value, we can use it for real things; money for food, and P-values for knowledge claims, publication, funding, and promotion. It does not matter if the P-value does not mean what people think it means; it becomes valuable because of what it buys.” Put another way, many of the criticisms of p-values and significance thresholds–that they oversimplify to the point of misrepresentation, that they falsely dichotomize between “significant” and “non-significant” results–are precisely why they have become so widely used: they provide a simplified and easy-to-digest (even if false and misleading) representation of research results. Researchers have decades of experience operating within a system that grants professional rewards to those using significance thresholds. Currently, I can envision only two plausible pathways away from such flawed practices. Either research fields need to converge on a viable consensus to replace the longstanding dominance of p-values and significance thresholds (rather than the range of conflicting opinions represented in the special issue). If such a simple consensus replacement is not possible, then scientific institutions (including journals, universities, and funding bodies) need to at least remove the incentives for engaging in flawed practice, so that a diverse range of less flawed practices can flourish.

Or, as Steven Goodman writes in his contribution to the special issue–”The use of P-values is a social phenomenon upon which many social rewards and penalties rest.”

Brittle anxious nonlinear incomprehensible

Most of her stuff is free – democratizing psychological knowledge –

Inside Out

The basic explanation is neither philosophical nor scientific, but sociologic; everyone uses them. It is the same reason we can use money. When everyone believes in something’s value, we can use it for real things; money for food, and P-values for knowledge claims, publication, funding, and promotion. It does not matter if the P-value does not mean what people think it means; it becomes valuable because of what it buys. Ted Porter, in his book “Trust in Numbers” has written that statistics are a refuge of objectivity for disciplines that are fighting other battles, usually against some form of claimed expertise of its members. Statistics in medicine pushed back against claims of physicians that they could predict what worked and in whom using physiologic or biologic reasoning: …the use of statistics tests has become obligatory in…scientific research. …they work mainly as social technologies, not as guides to private thinking. The advances of statistics in medicine must be understood as responses to problems of trust, which have been most acute in the context of regulatory and disciplinary confrontations. This, and not any inherently statistical character of clinical medicine, explain why inferential statistics entered medicine through therapeutics. (Porter 1995, p. 209) P-values are part of a rule-based structure that serves as a bulwark against claims of expertise untethered from empirical support. It can be changed, but we must respect the reason why the statistical procedures are there in the first place. This partly explains why there is so much resistance to Bayesian approaches, which are often viewed as a back-door way to reintroduce the subjectivity that conventional statistical methods were introduced to counter.