The nonsense math effect

Kimmo Eriksson*†

Abstract

Mathematics is a fundamental tool of research. Although potentially applicable in every discipline, the amount of training in mathematics that students typically receive varies greatly between different disciplines. In those disciplines where most researchers do not master mathematics, the use of mathematics may be held in too much awe. To demonstrate this I conducted an online experiment with 200 participants, all of which had experience of reading research reports and a postgraduate degree (in any subject). Participants were presented with the abstracts from two published papers (one in evolutionary anthropology and one in sociology). Based on these abstracts, participants were asked to judge the quality of the research. Either one or the other of the two abstracts was manipulated through the inclusion of an extra sentence taken from a completely unrelated paper and presenting an equation that made no sense in the context. The abstract that included the meaningless mathematics tended to be judged of higher quality. However, this "nonsense math effect" was not found among participants with degrees in mathematics, science, technology or medicine.

Keywords: unintelligibility, mathematics, humanities, social science, quality of research

1 Introduction

The background to this paper is my own subjective experience of a mid-career move from pure mathematics to interdisciplinary work in social science and cultural studies. In areas like sociology or evolutionary anthropology I found mathematics often to be used in ways that from my viewpoint were illegitimate, such as to make a point that would better be made with only simple logic, or to uncritically take properties of a mathematical model to be properties of the real world, or to include mathematics to make a paper look more impressive. In those areas in the social sciences and humanities in which it is broadly recognized that mathematics is a valuable tool, mathematical skills may still be generally too low for optimal use of this tool. There may be a lack of understanding of what mathematics can and should be used for and what it cannot and should not be used for. If mathematics is held in awe in an unhealthy way, its use is not subjected to sufficient levels of critical thinking.

Within the field of pure mathematics, academic writing is expected to be completely transparent to anyone knowledgeable about the mathematical concepts involved. A typical reviewer of a pure mathematics paper will not tolerate a sentence or paragraph for which the meaning is obscure. The same cannot be said about some other academic disciplines. The mathematical physicist

Alan Sokal famously composed a paper that was deliberately obscure and nonsensical; he submitted it to the journal Social Text where the editors accepted it for publication (Sokal 1996a, 1996b). This so called "Sokal hoax" demonstrated that there are experienced readers of research publications who will not necessarily react adversely to the fact that a text cannot really be made sense of. Indeed, as Sokal and Bricmont (1998) documented in a book following up on the hoax, obscurity is a hallmark of certain academic traditions associated with terms like postmodernism and relativism. However, these science-bashing genres were not responsible for the abuse of mathematics I referred to above; on the contrary, I found mathematics to be poorly used in fields where the tools of science are held in uncritically high regard. The acceptance and admiration of writing that actually does not make much sense may be found in both camps.

Of course, my personal impressions may be biased. Perhaps there is no widespread acceptance and admiration of obscure or nonsensical use of mathematics in the social sciences and humanities, or, if there is, perhaps it is just as widespread in areas like technology and the natural sciences where everyone learns how to use mathematics. In order to obtain some hard evidence, I conducted an experiment. In brief, I recruited participants with research experience from varying disciplines. Participants were presented with two abstracts (from published papers in good journals) and asked to judge the quality of the research presented in the abstracts. One of the two abstracts, randomly chosen, was manipulated by the addition of an extra sentence. This sentence was taken from a completely unrelated paper and presented an equation that made no sense in the context.

My prediction was that participants with a background in softer areas like humanities and social science would tend to judge the quality of research as higher when meaningless mathematics was included, whereas participants with a background in harder areas like mathematics, natural science or technology would not tend to be impressed with meaningless mathematics.

2　Method

The study demanded comparable sets of participants from different academic disciplines, all with experience of reading research reports. To find such participants I used Amazon's Mechanical Turk (mturk.com). This is an online labor market with many thousands of users of varying backgrounds who will do tasks for small monetary compensation. The usefulness of Mturk for online experiments is well documented (Paolacci et al., 2010). I advertised a task of judging the quality of research from abstracts, and asked for users with a postgraduate degree and experience of reading research reports. A fee of $0.50 was offered for approximately five minutes work.

2.1　Participants

Two hundred American adults (54% male, mean age 32 years) were recruited among users of the Mturk.

2.2　Questionnaire

Participants filled in an online questionnaire. It started with questions about their qualifications, including their postgraduate degree [either *Master's degree* (88%) or *PhD* (12%)]; the area of their degree [either *humanities or social science* (42%), *medicine* (8%), *mathematics, natural science or technology* (34%), or *other (e.g., education)* (16%)][1]; and their experience of reading research reports, such as journal articles, conference papers, anthology chapters or monographs [either *Have read less than 10 different reports* (12%), *Have read between 10 and 100 different reports* (54%), or *Have read more than 100 different reports* (34%)].

The questionnaire went on to describe that organizers of scientific conferences often ask for researchers to submit abstracts of the research they would like to present, and that based only on these short abstracts the highest quality research is to be selected. The current study was presented as an investigation of how readers of abstracts judge the quality of research.

Two abstracts were then presented. For each abstract, participants were asked to give their general judgment of the quality of the research. Responses were given on a scale from 0 (the very lowest quality) to 100 (the very highest quality). To mimic a typical procedure for judgment of submitted abstracts, participants were also asked to rate some other aspects of the abstract, such as its importance and how interesting it was.

The two abstracts were taken from real research papers, well-cited and published in very good journals. They were selected so that they would be generally understandable to non-specialists. The first abstract (referred to as "Foraging" below) described a study of whether food-sharing practices among a foraging tribe could be predicted from risk reduction and reciprocity (from Bliege Bird et al. 2002):

> Foragers who do not practice food storage might adapt to fluctuating food supplies by sharing surplus resources in times of plenty with the expectation of receiving in times of shortfall. In this paper, we derive a number of predictions from this perspective, which we term the risk reduction reciprocity (RRR) model, and test these with ethnographic data on foraging (fishing, shellfish collecting, and turtle hunting) among the Meriam (Torres Strait, Australia). While the size of a harvest strongly predicts that a portion will be shared beyond the household of the acquirer, the effects of key measures of foraging risk (e.g., failure rate) are comparatively weak: Harvests from high-risk hunt types are usually shared more often than those from low-risk hunt types in the same macropatch, but increases in risk overall do not accurately predict increases in the probability of sharing. In addition, free-riders (those who take shares but do not reciprocate) are not discriminated against, those who share more often and more generously do not predictably receive more, and most sharing relationships between households (over 80%) involve one-way flows.

The second abstract ("Incarceration") described a study of the consequences of incarceration for the employment outcomes of black and white job seekers (from Pager 2003):

> With over 2 million individuals currently incarcerated, and over half a million prisoners released each year, the large and growing number of men being processed through the criminal justice system raises important questions about the consequences of this massive institutional intervention. This article focuses on the consequences of incarceration for the employment outcomes of black and white job seekers.

---

[1]This division of research into four broad areas is used by the Swedish Research Council.

The present study adopts an experimental audit approach—in which matched pairs of individuals applied for real entry-level jobs—to formally test the degree to which a criminal record affects subsequent employment opportunities. The findings of this study reveal an important, and much underrecognized, mechanism of stratification. A criminal record presents a major barrier to employment, with important implications for racial disparities.

The questionnaire came in two versions, randomly assigned so that each version was given to 100 participants. The versions differed only in which of the two abstracts was presented in a manipulated version. The manipulation always consisted in the addition of the following sentence at the end of the abstract:

> A mathematical model ($T_{PP} = T_0 - fT_0d_f^2 - fT_Pd_f$) is developed to describe sequential effects.

This sentence was adapted from a completely unrelated paper on reaction times in choice experiments (Soetens et al., 1984). As none of the original abstracts mention any sequential effects or anything that the symbols in the equation could reasonably correspond to, the manipulation amounted to inclusion of meaningless mathematics.

2.3 Analysis

The design of the study was 2 abstracts to be rated [Foraging or Incarceration] × 2 conditions [math added to Foraging or to Incarceration] × 4 areas of degree [Math/science/technology, Medicine, Humanities/social science, Other (e.g., education)]. My main dependent variable is the *rating advantage of added math*, calculated as the rating of the manipulated abstract minus the rating of the non-manipulated abstract. My research questions were whether there is a positive rating advantage of added math (regardless of which abstract was manipulated), and whether the rating advantage was moderated by participants' area of degree. This was analyzed through an ANOVA of the rating advantage with condition and area of degree as factors.

3 Results

For participants from different areas, Table 1 presents descriptive statistics of the rating advantage of added math (conditions pooled). The table also includes the results of one-sample t-tests. The table indicates an overall positive

Table 1: Rating advantage of added math depending on the area of degree of participants.

Area of degree	N	Mean (SD) rating advantage of added math
Math, science, technology	69	−1.3 (19.2)
Medicine	16	3.0 (16.0)
Humanities, social science	84	6.6** (21.2)
Other, e.g., education	31	13.9** (23.3)
Total	200	4.7** (21.0)

* $p < .05$; ** $p < .01$.

effect of added math, which is moderated by the participants' area of degree such that effect is in evidence only outside the area of mathematics, science and technology.

The ANOVA confirmed that there was a rating advantage of added math, F(1,192) = 8.7, $p = .004$; there was no main effect of condition, F(1,192) = 1.8, $p = .18$, meaning that regardless of what abstract was manipulated the manipulation gave about the same rating advantage; there was a main effect of area of degree, F(3,192) = 4.2, $p = .006$, meaning that the rating advantage of added math differed between participants with degrees in different areas. There was no significant interaction between condition and area of degree, $p = .26$. The results of the ANOVA are robust to inclusion of experience of reading research reports as a covariate. (Indeed, the effect of the manipulation was at least as strong among those who had read more than 100 reports as among those who had read between 10 and 100 reports.)

Figure 1 illustrates the effect of the manipulation in terms of the percentage who rated the abstract with added math highest, excluding those who gave exactly equal ratings of both abstracts. The majorities for the areas "humanities and social science" and "other (e.g., education)" are statistically significant, ps<.05, binomial tests.

4 Discussion

Quality judgments of research are known to be influenced by cues outside the research itself, such as the prestige of the author and the author's institution (Peters & Ceci 1982; Garfunkel et al. 1994; Willer 2012). The experiment presented here demonstrated a related but arguably more problematic effect: Participants judged the quality of research as higher when the content included unintelligible elements, which arguably ought to detract from the quality.

Figure 1: The percentages of participants who gave the highest rating to the abstract with added math rather than the other abstract.

Area of degree
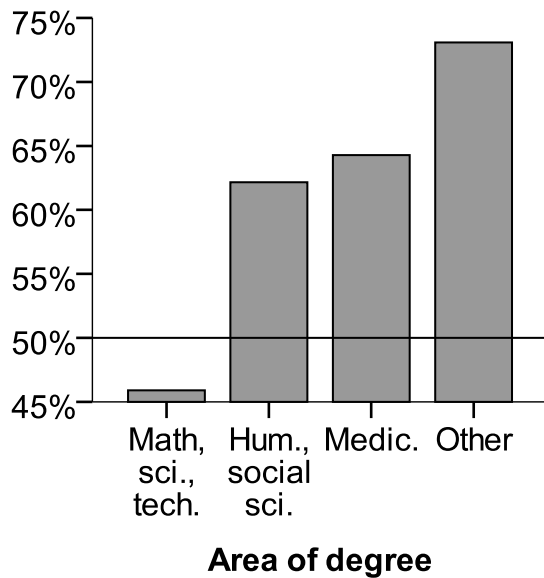
Specifically, the experimental results suggest a bias for nonsense math in judgments of quality of research. Further, this bias was only found among people with degrees from areas outside mathematics, science and technology. Presumably lack of mathematical skills renders difficult own critical evaluation of meaningless mathematics. Of course, this specific mechanism was not tested in the experiment. It is possible that the crucial difference between those with a degree in mathematics/science/technology and those with a degree in other areas is not training in mathematics but something else. Future research should specifically address whether, all other things equal, more training in mathematics decreases the bias. However, an indication that subject relevant expertise indeed removes the allure of irrelevant subject matter is found in a related study by Weisberg et al. (2008). They presented explanations of psychological phenomena and manipulated the presence of irrelevant neuroscience information. Non-experts judged explanations as more satisfactory if they contained the irrelevant information. In contrast, neuroscience experts rated explanations no more satisfactory, or even less so, if they included irrelevant neuroscience.

It may be that both mathematics and neuroscience are held in undeserved awe among nonexperts. It may also be that people always tend to become impressed by what they do not understand, irrespective of what field it represents—much in line with the "Guru effect" discussed by Sperber (2010). The scope of the phenomenon is a question for future research.

References

Bliege Bird, R. L, Bird, D. W., Kushnick, G., & Smith, E. A. (2002). Risk and reciprocity in Meriam food sharing. *Evolution and Human Behavior, 23*, 297– 321.

Garfunkel, J. M., Ulshen, M. H., Hamrick, H. J., & Lawson, E. E. (1994). Effect of institutional prestige on reviewers' recommendations and editorial decisions. *Journal of the American Medical Association, 272*, 137–138.

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology, 108*, 937–975.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences, 5*, 187–255.

Soetens, E., Deboeck, M., & Hueting, J. (1984). Automatic aftereffects in two-choice reaction time: A mathematical representation of some concepts. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 581–598.

Sokal, A. (1996a). Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity. *Social Text, 46/47*, 217–252.

Sokal, A. (1996b). A physicist experiments with cultural studies. *Lingua Franca (May/June)*, 62–64.

Sokal, A. D., & Bricmont, J. (1998). *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science.* New York: Picador.

Sperberl, D. (2010). The Guru effect. *Review of Philosophy and Psychology, 1*, 583–592.

Weisberg, D. S., Keil, F.C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470–477.

Willer, R. (2012). The effect of author's status on the evaluation of unintelligible texts. Working paper. University of California, Berkeley.