

At least one of the papers that James Damore cited is bogus, and that's completely unsurprising

James Thomas

Copyright 2017 James Thomas

News has just broken that a software developer at Google circulated a memo he'd written containing putatively scientific assertions to the effect that women are innately inferior to men in certain respects (but only "on average," so presumably we mustn't hate him too much) and recommendations to strive less for diversity in the workplace. So he was fired, and the press reacted predictably: those who call themselves liberals applauded the punishment of a heretic, and conservatives shook their heads at the intolerance of scientific "truth."

Which side is right? In the truest sense, neither is, because the problem is deeper than the irritation has let on, and its implications are more sinister. But here we focus on just one of the papers cited during the media storm, because even a nonscientist can show that it's bogus.

Bramble *et al.*

Before proceeding to the botched study, consider another one whose bogusness lies not in itself but in the way in which it's being cited. It's the one by Bramble *et al.*, entitled "Sex-Specific Effects of Testosterone on the Sexually Dimorphic Transcriptome and Epigenome of Embryonic Neural Stem/Progenitor Cells."

Sounds fancy, right? The abstract¹ sounds even fancier. But its essential message is simple: the genes of brain cells in mice are expressed differently at different levels of testosterone, so the male brain is different from the female one.

The problem with this claim—assuming it really is true of mice and assuming it can be generalized from mice to humans—is that it's *irrelevant*. We already know that there are sex differences in the mammalian brain. There are, in case you haven't noticed, sex differences in mammalian bodies. So, since the brains control the bodies, it would be no surprise if they too had sex differences.

The question, rather, is: how do sex differences make women inferior at software engineering or any of its related mental skills? The paper sheds no light on this question, and yet the essentialists—the "nature" side of this "nature-*vs.*-nurture" debate—cite the paper as if it somehow helps answer it.

Pasterski *et al.*

The botched paper, by Pasterski *et al.*, is available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.2005.00843.x/abstract>, and its abstract can be summarized simply. The experimenters took a group of children whose ages ranged from three to ten, some of whom were encouraged more than others to play with sex-typical toys, and some of whom have congenital adrenal hyperplasia (CAH), a disorder that virilizes a child prenatally, *i.e.*, makes the child more biologically male than normal (on boys it has little effect, but in girls it can make the genitalia ambiguous to the point of "pseudohermaphroditism").

¹<https://www.ncbi.nlm.nih.gov/pubmed/27845378>

The idea was to see if these three factors—CAH, encouragement, and biological sex—influence the child’s preference for sex-typical toys.

Before discussing the results, let’s recast the situation in the standard language of the scientific method, such as we find in high school or middle school science courses. It will shed light on the study’s mistakes. First, there are the three *independent variables* (“IVs”):

- *IV 1*: the sex of the child. This can take two values: male and female.
- *IV 2*: whether or not the child has CAH. This can take two values: the child either does or does not have CAH.
- *IV 3*: how much the child was encouraged to play with sex-typical toys. This can take two values: the child received either normal or extra encouragement to play sex-typically.

So, the eight possible combinations of the IVs’ values are:

1. A boy without CAH is given normal encouragement to play with sex-typical toys.
2. A boy without CAH is given extra encouragement.
3. Boy with CAH, normal encouragement.
4. Boy with CAH, extra encouragement.
5. Girl without CAH, normal encouragement.
6. Girl without CAH, extra encouragement.
7. Girl with CAH, normal encouragement.
8. Girl with CAH, extra encouragement.

Recall that the experimenters sought to measure how the child’s behavior (his or her toy preference) varied with changes in these IVs. This is the *dependent variable* (“DV”): whether or not the child’s preferred toys are sex-typical. So the DV can take two values: sex-typical or sex-atypical.

In order to perform the experiment thoroughly, *the DV has to be measured for all eight possible combinations of the IVs*. Interestingly, however, the experimenters failed to do so. Instead, their abstract reports the following:

- The boys were given normal encouragement, and those with CAH showed no more likelihood of sex-typical play.
- For girls:
 - Those without CAH were given normal encouragement and responded with a certain degree of sex-typical toy play.
 - Those with CAH were given extra encouragement and responded with *less* sex-typical play than the other girls did.

In other words, *the DV was measured only for combinations 1, 3, 5, and 8, above*. In the case of the boys (combinations 1 and 3), the IV 3 was *controlled* (kept fixed) while IV 2 was varied. For this reason, the finding of no change in the DV does indeed support the hypothesis that CAH does not affect a boy's behavior, as long as he is given no extra encouragement, and assuming that nothing more is wrong with the experiment (plenty more is indeed wrong with it, as discussed below).

But for the girls (combinations 5 and 8) the experimenters did *not* control the experiment. In fact, they botched it in a way that suggests a political agenda. Consider, for example, a researcher with reductionist bias. He or she might botch the experiment by measuring the DV only for combinations 5 and 7, or only for combinations 6 and 8. If the DV showed that the behavior varied with CAH, then the finding would seem to support preconceived notions that behavior (gender) is determined by biochemistry alone—the “nature” side of “nature-vs.-nurture.”

Conversely, a researcher with social constructivist bias might botch it by measuring the DV only for combinations 5 and 6, or only for combinations 7 and 8. Again, if the DV showed that behavior varied with encouragement, then the finding would seem to support preconceived notions that behavior is learned—this is the “nurture” side of the debate.

Now the correct way to perform the experiment, the way that tries to cancel the researchers' biases, is to measure the DV for all four combinations—5, 6, 7, and 8—and to try to make sense of the results. If they came out, say, as follows

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
5	yes	normal	typical
6	yes	extra	typical
7	no	normal	typical
8	no	extra	typical

then one would conclude that neither CAH nor encouragement determines behavior. The conclusion, if behavior is indeed determined, is that it must be determined by something else; other variables would need to be sought as candidate determiners.

If, on the other hand, the results came out as

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
5	yes	normal	atypical
6	yes	extra	atypical
7	no	normal	typical
8	no	extra	typical

then they would support the hypothesis that behavior is determined by CAH and not by encouragement. “Nature” would in this case win out over “nurture.”

Now consider this unusual scenario:

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
5	yes	normal	typical
6	yes	extra	typical
7	no	normal	atypical
8	no	extra	atypical

It would be a surprise, because although it would support reductionist predictions, it would also support the strange hypothesis that absence of CAH—*i.e.*, normal sexual development—produces atypical behavior. Such a result should immediately arouse the suspicion that a mistake has been made, perhaps in data collection. Whether or not that turned out to be the case, the experiment should of course be repeated many

times, to be sure that the results are not a fluke. This scenario is included here just to keep your mind alert to “edge cases” and to give you a larger sense of the way of thinking.

Finally, if the results came out like so:

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
5	yes	normal	atypical
6	yes	extra	typical
7	no	normal	atypical
8	no	extra	typical

then they would support the hypothesis that encouragement, and not CAH, determines behavior. “Nurture” would win out over “nature.” But it’s also important to appreciate that the same hypothesis would also be supported if the results were “inverted,” like so:

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
5	yes	normal	typical
6	yes	extra	atypical
7	no	normal	typical
8	no	extra	atypical

In this case, not only would we find support for determination by the environment alone (“nurture”), but we would call the results counterintuitive, because they would show that extra encouragement in fact *discourages* typical behavior. An explanation would therefore be in order. One possibility is that the children in the sample are rebellious, are somehow in the habit of defying their parents’ expectations. It would of course then be reasonable to repeat the experiment with a completely different sample of children. If the same results obtained, and if they keep obtaining with other samples of children, then we might suspect that something in the culture being studied makes children systematically defiant of their parents’ wishes.

Another possibility, which anticipates the deeper critique of this whole line of “research” below, is that the children are not rebelling, but are tuning in to something *else*, an independent variable other than IV 1, IV 2, or IV 3. Such a possibility would mean that, rather like the first example above, the experiment is searching for the wrong effect.

And now for what Pasterski *et al.* actually did. To be sure, they didn’t botch the experiment in the naïvely reductionist way, and they didn’t botch it in the naïvely constructivist way. The fact that they didn’t test all possible combinations of the IVs (5 through 8, inclusive) shows that they did botch it. But they did so in a telling way, for they measured the DV only for combinations 6 and 7, like so:

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
6	yes	extra	atypical
7	no	normal	typical

The appearance in this table of all four values of IV 2 and IV 3 makes it look as though all four combinations have been tested. But of course they haven’t. Moreover, the speciousness is *convenient to reductionists*, because the first row of this table (combination 6) makes it look as though the experimenters, by giving extra encouragement for typical play in CAH girls, went to the trouble of “cancelling out” any possible environmental influences, whereas they have not in fact done so. To have done so, they would have measured the value of the DV for combination 8 and shown it to us. If the results had turned out like so

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
6	yes	extra	atypical
7	no	normal	typical
8	no	extra	atypical

then the reductionists would lose, because it would show that something about the encouragement itself is causing atypicality. If, on the other hand, the results turned out like so:

combination	CAH (IV 2)	encouragement (IV 3)	behavior (DV)
6	yes	extra	atypical
7	no	normal	typical
8	no	extra	typical

then the reductionists would be supported, though of course they would need to show, by much repetition of the experiment, that the effect is reproducible—that it’s real.

Many scientists—more importantly, many *nonscientists* whose common sense hasn’t succumbed to blind trust in so-called “experts”—would of course agree with the foregoing assessment of the study’s errors. But the sad fact is that we live in an age in which even such people miss the much deeper problems with the study, if not with the whole field. To see what I mean, consider again the three independent variables. IV 1 is biological sex, which the study models as having two possible values, male and female. This is hardly a problem, since in the vast majority of people the twenty-third chromosome pair is either XX (“female”) or XY (“male”).

But IV 2, which indicates whether or not the child has CAH, is not quite so simple. Consider this quote from Wikipedia:²

Further variability is introduced by the degree of enzyme inefficiency produced by the specific alleles each patient has. Some alleles result in more severe degrees of enzyme inefficiency. In general, severe degrees of inefficiency produce changes in the fetus and problems in prenatal or perinatal life. Milder degrees of inefficiency are usually associated with excessive or deficient sex hormone effects in childhood or adolescence, while the mildest forms of CAH interfere with ovulation and fertility in adults.

In plainer language: there are different degrees of CAH, and the disorder manifests itself in different ways in different people. Thus the meaning of CAH varies, both qualitatively and quantitatively. And yet the study treats it as a single two-valued variable. Such a model *coarsens* reality, from a multi-dimensional continuum (if that itself is not too scientific an oversimplification) to a binary choice. Note that there is no such flaw in modeling, say, the concentration of a solute, or the location of a particle with respect to some coordinate system. Real sciences, such as physics and chemistry, suffer no ambiguity of definition in their variables.

But nevermind, because these problems are nothing compared to the problems with IV 3, which *should*, in a sane society, take people’s breath away. Recall that it is supposed to measure the amount of encouragement given to the children. Working our way from smaller to bigger problems, let’s begin by considering that word “measure.” How, pray tell, can this be done? The experimenters could only have done something equivalent to putting parents and children in an observation room filled with toys, and counting the number of times they heard the parents say “Good girl!” But this is fatuously stupid. For starters, the parents and children are also interacting *outside* the “laboratory,” so how can the count be reliable? What if the parent says “Good girl!” twice in a row, in one breath; does that count as one instance of encouragement or two? What if the second time swallows up the words, or mutters them softly? What about “feedback” that comes from persons *other* than the parents? Since these children’s ages range from three to ten, they are constantly being exposed to messages from school, from other family members, from friends and neighbors, and, perhaps most overwhelmingly, from the media. What kind of idiocy is it to think that this nonstop barrage of

²https://en.wikipedia.org/wiki/Congenital_adrenal_hyperplasia, as of August 20, 2017.

influence can be magically waved away by the fiat of a “researcher,” and replaced by a thoughtless count of the number of times a formula is uttered?

And speaking of age, has nobody noticed that it is in fact a *fourth* independent variable, and is being left *uncontrolled*? What makes this lapse even more ironic is that age is perhaps the most “ontologically stable” variable: unlike the others (even biological sex!) there is absolutely no ambiguity in its definition, no difficulty in measuring, with fabulous precision, how old a child is. (To belabor the point further: even if the child’s age were to suffer relativistic twin-paradox modifications owing to super-fast space travel, it would still be well defined in principle and easy to measure in practice.) So why is nobody shocked by the fact that it wasn’t controlled?

The authors also presume to know how to classify toys as “sex-typical” or not. It’s a good thing they’re not doing the study in Sweden, where I hear it is normal for boys to knit. Or in Scotland, where people love to see them wear kilts. One can only wonder how to decide that a given thing is “sex-typical” for the poor brain that has been partly exposed to American culture, partly to Swedish, and partly to Scottish.

But all of these glaring problems are *still* nothing when we consider one of the deepest and most basic facts about communication—namely, that its real meaning lies not in what is literally said, but in what is *not* literally said. As this same corps of so-called experts has gotten our society scientistically to recite, “Communication is 90% nonverbal.” (For now, let’s overlook the fatuity of modeling “communication” as a simple quantity, let alone of assigning a percentage to its nonverbal part.) Take any slur, for example—racial, ethnic, or sexual. Say the word the wrong way and you could incite a riot. But say it under the right circumstances and you could seduce a person into intimacy, might even excite sexual desire. Indeed, if the minimum age of this sample weren’t three years, one could easily have argued that the subjects must be even *more* sensitive to “nonverbal communication,” for the simple reason that they haven’t yet learned the language. No matter: we have the example of dogs. Speak to one in any language, and it will, all things being equal, respond the same way to the same *tone*, to the same *body language*. In fact, if we really need an “expert” to confirm the obvious fact that *nothing* needs to be actually said in order for communication to take place, then we can find it in remind us that communication doesn’t need *anything* to be said, then we can find it in Hastings’ book on child sexual abuse [Has94], in which she notes that a breastfeeding mother can tell that the baby can sense whether she is sexually aroused by the sucking.

So if our brilliant experimenters do indeed hear a parent speak *words* of encouragement to their children, how can they know that it will have the effect they expect? The very fact that a parent knows that the child is sexually abnormal means that the parent is more likely to be *anxious* about making the child “gender-typical.” Are we really supposed to believe that the child will not tune in to such anxiety, in the parent’s voice, face, or body? Are we really supposed to believe that these “variables,” which are not only unaccounted for but fundamentally *unaccountable*, will not produce counterintuitive effects? And even if such foolery is to be countenanced, has no one noticed that the experiment is botched by the very fact that it isn’t *blind*, that the parents actually *know* which kids have CAH and which don’t?

The larger question is: why does the public need to be reminded of such truisms? Why does it feel that a special license is needed to call pseudoscientists out on their nonsense? Don’t be surprised if the so-called “experts” counter, with a straight face, that such effects cannot be real because they cannot be measured. Or, alternatively, they may acknowledge that the effect is real, thank us for pointing it out, and come back saying that they’ve increased the number of independent variables by including in the experiment a spectral analysis of the parents’ voices, to somehow detect whether or not the praise is sincere. Throw in a mention or two of machine-learning algorithms and the gullible public will drool with admiration. In vain is it pointed out that the thing that counts here is the child’s *experience* of the parent’s voice, let alone all the other “variables.” And experience, by its very nature, is not something that can be measured objectively.

Such simple, common-sense exposures of the radical idiocy of this “science” shows that we are light-years away from the straightforwardness and well-definedness of, say, counting the number of electrons in an atom, or measuring the distance to a planet. This is not science, but derangement and delusion.

If this business now has you depressed, then prepare for worse. For starters, the authors' affiliations show them to belong to highly regarded institutions—UCLA, University College London, City University in London, *etc.*—so are we to suppose that work at so-called “lesser” schools and hospitals is even shoddier? And considering that the paper was actually published, it presumably went through a process of peer review, of scrutiny by a handful of anonymous colleagues drawn from a pool of “behavioral scientists” stretching across the globe. That they approved the paper for publication raises obvious questions about *their* competence too.

But even this is not surprising, because the habit of radical error and incompetence in these pseudoscientific departments was already noted half a century ago, by a number of scientists who went to the trouble of raising alarms about it. They include people such as Ruth Hubbard, Richard Lewontin, Leon Kamin, Steven Rose, Stephen Jay Gould, William Byne, and Richard Feynman, no less. As for the problems of peer review, Michael Crichton was so disturbed by its corruptions that he took the trouble to speak out about it. Of course, all of these alarms have been systematically ignored. (Why? Think!)

And the scandal doesn't really stop there. It involves you and me as well. To understand what I mean, consider that the principles of the scientific method are based on nothing but common sense. They don't require any calculus, any algebra, any abstruse terminology (unless we call the word “variable” abstruse), or any arcane mathematics. They require no special training at all. The only skill, apart from high-school biology, that we used to expose the mistakes of Pasterski *et al.* is elementary *logic*, which might better be called simple common sense. And yet we grant funds and influence to these quacks, look away, trust that they know what they're doing, trust their internal processes of quality control, trust that the media would never hype up results that are wrong or stupid, and, as if all that were not bad enough, allow jurists and legislators to embed their fatuities in the law. It's true that those without higher training can't evaluate the work done in the *real* and advanced sciences; subatomic particles, for example, are not part of our everyday lives. But *behavior* has been observed, engaged in, reflected on, and talked about, systematically commented on, by everyone and for many millenia. How did the twentieth century persuade us to relinquish that authority to pseudoscientists?

The question turns out to be too deep for this short essay, so the interested reader will need to study some cultural history and criticism, beginning perhaps with Jacques Barzun's *Science: The Glorious Entertainment*. Those wanting a sense of the epidemic incompetence in the so-called behavioral sciences can read Feynman's short graduation speech, “Cargo Cult Science.” And Jeffrey Masson, particularly in his *Against Therapy*, mounts careful attacks on the constellation of shams known as “psychotherapy.”

References

- [Has94] Anne Stirling Hastings. *From Generation to Generation. Understanding Sexual Attraction to Children*. The Printed Voice, 1994.