

At least one of the papers that James Damore cited is bogus

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 those on the right 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.

The 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"). 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.

And for complete clarity, let's spell out the eight possible combinations of the IVs' values:

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 (for sex-typical play, of course) 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 experiment was *controlled*, in that IV 3 was 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, but there isn't enough space here to treat those more fundamental errors).

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 behavior varied with CAH, then the finding would support preconceived notions that behavior (gender) is determined by biochemistry alone—the “nature” side of “nature-vs.-nurture.”

Conversely, a researcher with social constructionist 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 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, since we're now restricting our attention to the girls—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 next implication would be that, if behavior is indeed determined, then it must be determined by something else; other variables would need to be sought as candidate determinants.

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

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.”

Before moving to the next scenario, it’s worth considering this unusual one:

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 suspicions of a mistake 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 were not a fluke. Again, this unusual scenario is included here only 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, determine 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" (a critique I don't have space here to mount) 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, *i.e.*, that it is real.

It turns out, however, that these latter points are moot, because even if the authors had bothered to obtain the missing data, the experiment would *still* be invalid, in a much more fundamental way. The reasons for this radical failure don't fit in a paper of this size, so they'll have to wait for a later discussion.

But for now, we note that the fact that the authors did not show us the result for combination 8 means that the experiment is even more invalid, and the reasonable person concludes that they are incompetent. But that they should have done it in this specious way; that these are the same flaws about which eminent scientists from Richard Feynman to Stephen Jay Gould have for decades been sounding the alarm; that they should persist until today, in a climate that places such high political stakes on the outcome; that big pharma corporations will lose billions of dollars in current and future profits if they fail to make the public believe in reductionism; all these circumstances make the reasonable person wonder if something more than simple incompetence is at work here.