Inferences of Cause-and-Effect from Correlational Data

Kate Morrison and Hali Bebault
Department of Psychology, University of Wisconsin-Eau Claire
Dr. April Bleske-Rechek, Faculty Mentor

Overview

Most college students can easily complete the phrase, "Correlation does not imply…" with the word "causation." But how well do they actually apply this reasoning? History is filled with examples of humans' inherent tendency to infer cause-and-effect from a mere association (e.g., the long-held inference, from correlational data on authoritative parenting and children's good behavior, that authoritative parenting leads to children's good behavior). In the abstract, the implication is that when researchers state that "Variable X" and "Variable Y" are correlated, people may mistakenly infer that "Variable X" causes "Variable Y."

We hypothesized that the conflation of correlation with causation is common. We speculated that the order in which variables are presented has an influence on which variable is assumed to be the cause and which the effect; and that the influence of variable order is most robust when correlations are presented both visually and in text. To test these hypotheses, we generated five correlational scenarios and presented each scenario to students in one of six different versions. The versions were created by combining two independent variables: (1) the order of X and Y (X before Y, Y before X); and (2) the form of presentation (text only, scatterplot only, text and scatterplot combined). After reading, viewing, or reading and viewing each scenario, participants restated the findings in their own words.

Method

A total of 230 students (85M, 145F) across the UWEC campus participated in this study as part of an in-class activity. Students came from a variety of majors; 51% were first-year students, 23% were second-year students, 16% were in their third year, and 10% were in their fourth year or beyond.

Each participant responded to five sets of findings. The five scenarios were all presented in the same format for each participant, depending on the condition s/he had been randomly assigned to. The five sets of findings involved the following pairs of variables:

Women's self-esteem ("x")
Women's number of sex partners ("y")

Children's IQ ("x")
Amount of spanking from parents ("y")

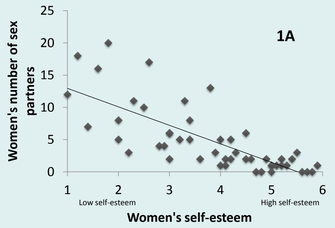
Level of confidence ("x")
Number of friends ("y")

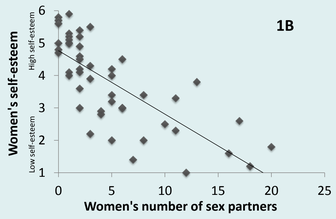
Children's level of disobedience ("x")
Parents' marital conflict ("y")

Couples' sexual satisfaction ("x")
Couples' emotional intimacy ("y")

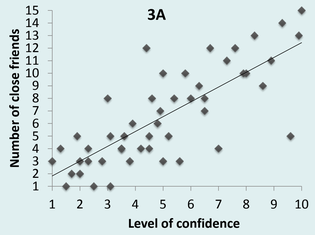
For half of participants, the variables were presented in x-then-y order; for half, the variables were presented in y-then-x order. Each of these two groups of participants was split into three subgroups. One group received the findings in text format only, such as: "Researchers have documented that women's self-esteem correlates negatively with their number of sex partners. Restate this finding in your own words." A second group received the findings in the form of a scatter plot only: "Researchers have documented the correlation shown above. Restate this finding in your own words." A third group received the findings in a combined format: "Researchers have documented that women's self-esteem correlates negatively with their number of sex partners, as shown above. Restate this finding in your own words."

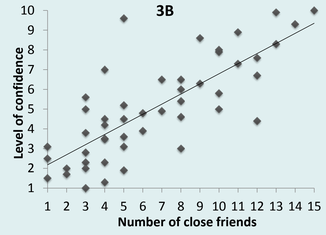
Scatterplots that participants viewed

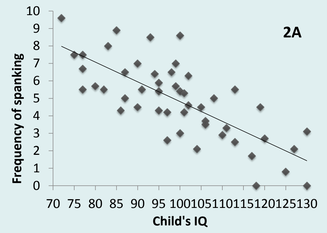


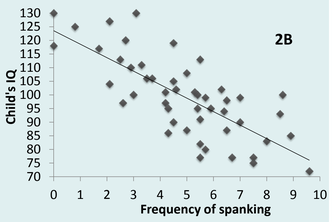


Each participant received either scatterplot A or B for each of the five findings. The format of presentation (in text format or scatterplot format) did not influence participants' restatements of the findings; however, scatterplots betrayed some confusions. First, negative correlations were interpreted as a lack of an association by several participants. Second, some participants misinterpreted the magnitude of the correlations, describing them as "weak" and "sporadic," when in reality, all associations were very strong (1A & 1B, *r* = -.75; 2A & 2B, *r* = -.75; and 3A & 3B, *r* = .78). Scatterplots for scenarios 4 and 5 are not shown here, but were also strong (*r* = .87 and .77).



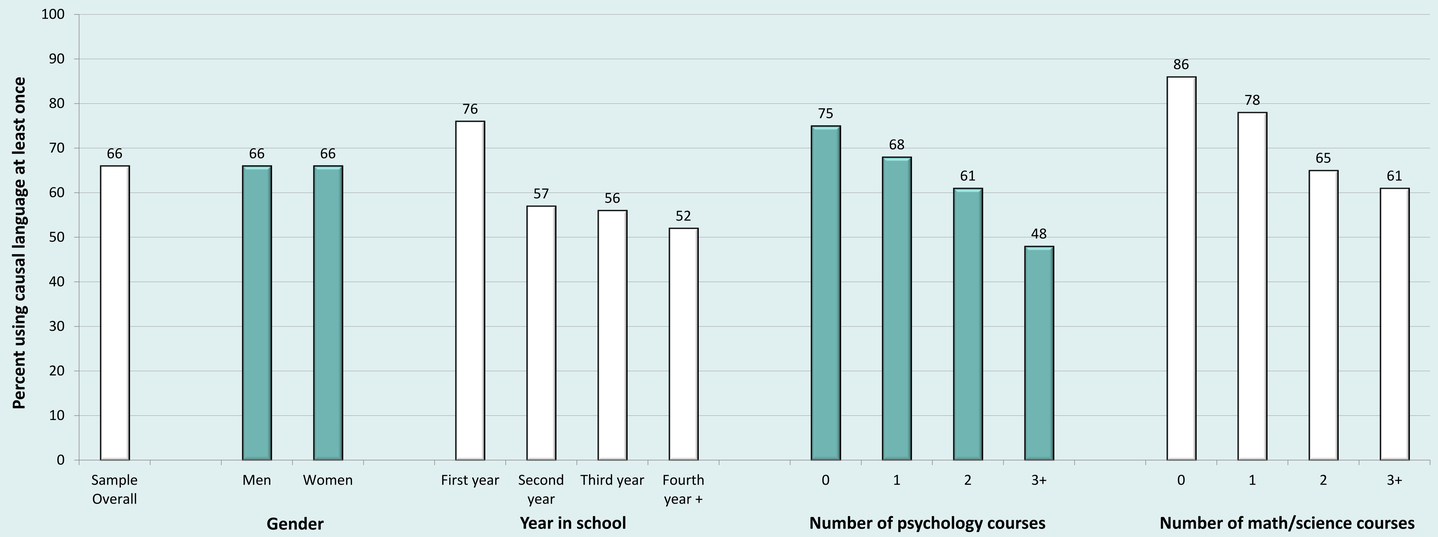






Results

Participants' use of causal language by gender, year in college, and exposure to scientific thinking



Our first hypothesis was supported: Participants frequently conflated correlation with causation. Overall, 66% of the sample used clear causal language in one or more of their restatements (see examples below). Men and women did not differ in their likelihood of using causal language, but younger students and students who had less exposure to psychology coursework and math-science coursework were more likely to use causal language. However, nearly one-half (48%) of students with 3 or more psychology classes still conflated correlation with causation.

Order of variables in participants' restatements

The order in which participants received the variables did not influence their restatements. Instead, the majority of participants used the same order, regardless of what order they had been exposed to. For example, in scenario 4, 74% of participants listed parents' marital conflict before children's level of disobedience in their restatements, regardless of the order the two variables had been presented in. Thus, even though some participants' statements weren't clearly causal, students had a clear preference for which variable came first. The fact that the majority of participants placed marital conflict before children's level of disobedience suggests that participants held preconceived notions that marital conflict precedes – or causes -- children's disobedience. In other words, many students used language that wasn't clearly causal, but consistency in their restatements betrayed cause-and-effect inferences.

Scenario	Variable X	Variable Y	% using causal language	Examples of causal language	% using non-causal, "fuzzy" language	Examples of non-causal, "fuzzy" language that may reflect causal thinking (written in order used more often)	% of sample who used the same order in their restatement (Chance = 50%)
1	Women's self-esteem	Women's number of sex partners	26.5%	When women have sex with an increasing number of sex partners it negatively affects their self-esteem.	60.4%	As women's number of sex partners increases, their self-esteem decreases.	65.5% implied that sex partners precede self-esteem
2	Children's IQ	Frequency of spanking	33.9%	The less a child is spanked, the higher IQ he/she will have.	52.6%	Children who get spanked more have lower IQs.	70.0% implied that spanking precedes low IQ
3	Level of confidence	Number of friends	33.5%	The more friends you have, the higher your confidence will be.	54.8%	More confidence = more friends.	66.5% implied that confidence precedes friends
4	Children's level of disobedience	Parents' marital conflict	39.6%	Higher marital conflict leads to higher levels of child disobedience.	47.0%	Children who live in a conflicted household are more disobedient.	74.4% implied that marital conflict precedes disobedience
5	Sexual satisfaction	Emotional intimacy	29.1%	If couples are more emotionally intimate, they will be more sexually satisfied.	58.3%	Couples have a higher level of sexual satisfaction when their level of emotional intimacy is higher.	72.1% implied that emotional intimacy precedes sexual satisfaction

Discussion

In many psychology classes, instructors demonstrate that correlation does not imply causation through the well known example of the positive association between *ice cream sales* and *murder rate*. In this example students have an easy time seeing the absurdity in inferring that ice cream sales have an effect on the murder rate or that the murder rate has an effect on ice cream sales; they easily generate third variables, like warm weather, that explain the association. In most questions that social scientists are interested in, however, the limitations of correlational data are not so easy to recognize. In the current study, we presented students with five correlational scenarios in which either variable could be causal and in which it is possible to generate third variables that could underlie the associations. We found that students did infer causation from data described as correlational, likely because they had preconceived notions about the issues involved. As Stanovich (2010) notes, "when the causal link seems obvious to us, when we have a strong preexisting bias, or when our interpretations become dominated by our theoretical orientation, it is tempting to treat correlations as evidence of causation." The current study showed evidence that the majority of students do this even in a circumstance where, if anything, they should have been primed not to (i.e., they were in a classroom, the study was introduced as being about "interpretations of data," and the scenarios we gave were worded and displayed in a scientific format).

We think the danger of inferring causation from correlation is bigger than some people might think. Take, for example, the scenario we gave participants in which children's IQ is negatively correlated with the frequency with which they are spanked. Some people might argue that it makes sense to assume that spanking precedes children's IQ, as 70% of our sample did. However, there are several reasons why common sense is not good enough. First, researchers cannot manipulate either IQ or spanking to test their causal effects. Second, it may not occur to people, but it is just as plausible that low IQ can elicit spanking or that a third variable, such as financial duress or parental IQ, could explain both low IQ and spanking. Third, history is full of examples in which the public has invested in mistaken causal, but common sense, interpretations of correlation data. The inference that erratic eye movements cause reading disabilities from the mere association between the two variables and the inference that poverty causes schizophrenia are two cases in point.

Future research in our lab will focus on further understanding how people interpret associations between two variables under different circumstances. Specifically, we plan to provide specific information about different research designs and ask people to tell us which of a variety of possible conclusions they could infer from the findings.

Acknowledgments

We thank ORSP for supporting this research and faculty across campus for allowing us to collect data in their classrooms: Monty Ernst, John Grauer, Valerie Guyant, BJ Hollars, Traci Thomas-Card, Catya von Karolyi, and Jeffrey Wright.