The Taboo Against Explicit Causal Inference in Nonexperimental Psychology

Michael P. Grosz\textsuperscript{a}, Julia M. Rohrer\textsuperscript{b,c}, and Felix Thoemmes\textsuperscript{d}

\textsuperscript{a} University of Münster, Department of Psychology, Fliednerstraße 21 (Pavillon 1), 48149 Münster, Germany. E-mail: michael.grosz@uni-muenster.de

\textsuperscript{b} Max Planck Institute for Human Development, Berlin, Germany

\textsuperscript{c} University of Leipzig, Department of Psychology, Leipzig, Germany. E-mail: julia.rohrer@uni-leipzig.de

\textsuperscript{d} Cornell University, Department of Human Development, Ithaca, New York. E-Mail: fjt36@cornell.edu

\textit{Perspectives on Psychological Science, in press}

This is an unedited manuscript accepted for publication at Perspectives on Psychological Science. The manuscript will undergo copyediting, typesetting, and review of resulting proof before it is published in its final form.
Author Note

We would like to thank Mitja D. Back, Steffen Nestler, and Ivan Flis for stimulating discussions and suggestions, Jane Zagorski for proofreading, and Hanna Krabbe and Christian Blötner for their help in preparing the manuscript.

Correspondence concerning this manuscript should be addressed to Michael P. Grosz. E-mail: michael.grosz@uni-muenster.de; Phone: +49 (2 51) 83 - 34279; Mail: Department of Psychology, University of Münster, Fliednerstraße 21 (Pavillon 1), 48149 Münster, Germany.
Abstract

Causal inference is a central goal of research. However, most psychologists refrain from explicitly addressing causal research questions and avoid drawing causal inference on the basis of nonexperimental evidence. We argue that this taboo against causal inference in nonexperimental psychology impairs study design and data analysis, holds back cumulative research, leads to a disconnect between original findings and how they are interpreted in subsequent work, and limits the relevance of nonexperimental psychology for policy making. At the same time, the taboo does not prevent researchers from interpreting findings as causal effects—the inference is simply made implicitly, and assumptions remain unarticulated. Thus, we recommend that nonexperimental psychologists begin to talk openly about causal assumptions and causal effects. Only then can researchers take advantage of recent methodological advances in causal reasoning and analysis and develop a solid understanding of the underlying causal mechanisms that can inform future research, theory, and policy makers.

Keywords: causal inference, observational studies, nonexperimental, instrumental variable estimation
The Taboo Against Explicit Causal Inference in Nonexperimental Psychology

Correlation does not imply causation. This truism justifiably reminds researchers that they should not carelessly draw causal conclusions on the basis of nonexperimental evidence. However, instead of motivating psychologists to exercise due diligence and face the challenges of causal inference, it seems to have resulted in a widespread taboo against explicit causal inference in nonexperimental settings. This taboo has resulted in a dilemma in some fields of psychology. On the one hand, causal relationships are of central interest; on the other hand, they are “forbidden” when experiments are unfeasible or unethical. As a result, one might expect nonexperimental researchers to limit themselves to descriptive or predictive research questions. But nonexperimental researchers do not actually avoid asking causal research questions or drawing causal conclusions; it simply happens implicitly, opaquely, and without an articulation of the underlying assumptions.

Here, we are going to argue that the taboo against explicit causal inference hinders nonexperimental psychology (for a similar argument in epidemiology, see Hernán, 2018a). It impairs study design and data analysis. It slows down the pace at which our understanding of underlying causal mechanisms grows. It leads to a disconnect between original studies and how they are integrated into subsequent work, and it limits the usefulness of nonexperimental psychology for policy making. We will elaborate on each of these points and suggest how nonexperimental psychologists can integrate causality into their work in a more productive manner.

Manifestation of the Taboo

To illustrate the taboo against causal inference, we annotated text passages from four nonexperimental articles in Table 1. In all four articles, causal language and causal inference seem to be intended but are not expressed in explicit and straightforward ways, which we interpret as signs of the taboo. The absence of explicit causal language thus obscured the research goals of the studies. The researchers ended up sending mixed messages. Some parts
of the articles read as if the entire endeavor was noncausal; yet other parts only make sense in the context of trying to answer a causal research question. For example, two of the four examples (Examples 1 and 3) included control variables to rule out confounding effects and estimate the unique effect of the independent variable. Although there might be instances in which third variable adjustment is useful for descriptive purposes, the adjustment in these two studies suggests that their goal was not merely to describe or predict. If description was the ultimate goal, then an unadjusted estimate could have been reported, potentially with some insightful graphical display. At the same time, prediction did not seem to be the goal either given that attention was paid to the coefficients of particular predictors, rather than to the overall accuracy and cross-validation of the prediction. If prediction was the ultimate goal, then techniques from the field of machine learning would have been more appropriate (Yarkoni & Westfall, 2017). Hence, it is plausible to assume that the two examples that used control variables and even the two examples without control variables (Examples 2 and 4) intended to identify causal links even though the results of the studies were interpreted with vague causal language (e.g., “predict,” “effect”) rather than explicit causal language (e.g., “causes,” “causal effect”; for a similar argument, see Rutter, 2007). We think that description and prediction are worthwhile research goals. Yet, we believe they should be pursued for their own sake rather than serving as a cover for implicit causal questions and conclusions.

To be clear, we do not intend to criticize the authors or the quality of these four articles; our intention is instead to criticize the norms regarding causal inference that these authors and most other nonexperimental psychologists adhere to. These norms permeate many aspects of psychological science, from the education of psychological researchers (e.g., causal effect estimation based on nonexperimental evidence plays only a minor role in the methods and statistics curricula in psychology) to the review process at scientific journals (e.g., author guidelines, editors, and reviewers asking for the removal of causal language).

**Reasons Behind the Taboo**
Why do psychologists think that it is legitimate to make explicit causal inferences on the basis of experimental evidence but not on the basis of nonexperimental evidence? Imagine that we wish to study the effect of a new therapy on the recovery of depressed people. In a purely observational study, we may observe that clients improve when undergoing treatment relative to the nontreated. Yet, it is likely that people who underwent treatment differed from nontreated people with respect to background factors that determine recovery (e.g., age, education, financial resources, social support). Thus, we cannot directly infer that the therapy worked—changes in recovery rates might be due to the treatment or due to those other factors. In an experimental study, the randomized assignment to the treatment (i.e., therapy) and control conditions intends to eliminate the causal link between the background factors and choice of treatment. Hence, the background factors cannot serve as an alternative explanation of a higher rate of recovery in the treatment group if the randomization was successful (e.g., Pearl, 2009).

However, the problems of multicausality that render nonexperimental evidence weak and potentially nondiagnostic are to some extent present in experimental research with randomized groups as well. This is because many treatment or experimental manipulations will affect not only the independent variable they are intended to affect. Even an obvious physical manipulation such as stimulus presentation time can have many causal effects at different levels of aggregation (e.g., by inducing time pressure or stress, undermining self-efficacy, inducing distinct cognitive strategies), and identifying the relevant mechanism might be challenging (Bullock, Green, & Ha, 2010). Hence, causal inference always goes beyond what is observed, and it always rests on assumptions (e.g., Waldmann, Hagmayer, & Blaisdell, 2006). Some philosophers have even argued that it is a top-down rather than a bottom-up endeavor that involves a priori world knowledge (e.g., Kant, 1781/2002).¹

¹ We would like to thank Klaus Fiedler for bringing up this point.
together, even though experimental designs are the gold standard for blocking the effects of background factors, causal inferences are speculative inferences, regardless of whether the study is conventionally classified as nonexperimental or experimental.

**Consequences of the Taboo Against Explicit Causal Inference**

**Impairment of Study Design and Data Analysis**

The ambiguity in the goals of nonexperimental studies (see Table 1) brings about a distinct lack of careful and explicit causal reasoning in study design and data analysis. Nonexperimental psychologists will usually have a coarse mental representation of the causal network in which their variables of interest are embedded. That is, they usually have some assumptions about the causes and consequences of the variables they are studying and about the causal mechanisms and mediating processes that lead from the independent variable(s) to the dependent variable(s). Yet, these assumptions about the underlying causal network are hardly ever spelled out explicitly. For example, many nonexperimental psychologists do not explicitly justify why they include certain control variables, and hardly any of them use formalized frameworks developed to support causal reasoning such as the potential outcome framework (e.g., Morgan & Winship, 2015; Rubin, 2005) or directed acyclic graphs (e.g., Pearl, 2009). As a consequence of this unstructured approach, researchers may forget to assess and control important confounding variables; or they may erroneously control for mediators and collider variables, hence introducing bias (e.g., Elwert & Winship, 2014; Foster, 2010a; Rohrer, 2018). This state of affairs was bemoaned by Foster (2010a) after editing the journal *Developmental Psychology* for 5 years: “Currently, developmentalists conduct complex analyses that are not useful in pursuing either aim: The analyses are too complex to produce good description, and the complexity is not employed in a manner that facilitates causal inference” (p. 1760).

Furthermore, the causal assumptions encoded in structural equation models are often ignored or at least not discussed openly. For instance, by setting a coefficient to zero in a
structural equation model, one is assuming that one variable does not have a causal effect on another variable. But structural equation models are frequently used in nonexperimental research without any explicit discussion or justification of such causal assumptions. This is problematic because the credibility of a structural equation model depends on the credibility of its causal assumptions (e.g., Bollen & Pearl, 2013).

**The Taboo Holds Back Cumulative Research**

A further consequence of the reluctance to explicitly talk about causality is that our understanding of the underlying causal mechanisms progresses at a slow pace, *if at all.* This issue has been highlighted in the field of personality research, which, due to the nature of its research subject, relies heavily on nonexperimental data:

> During the past 50 years, personality psychology has made considerable progress concerning personality description, and prediction of and by personality. In contrast, explanation of personality development and personality effects has lagged far behind. In the coming decades, much more inspiration and transpiration are needed to change this unsatisfactory situation. (Asendorpf et al., 2016, p. 305)

We believe it is currently difficult for fields strongly characterized by nonexperimental research to accumulate causal knowledge because most previous studies have not explicitly stated the causal link they have identified or the assumptions under which this link should hold. These assumptions can often only be reconstructed indirectly on the basis of the analyses the authors chose to apply. For example, controlling for a third variable implies that it is understood as a confounder rather than as a mediator of the effect of interest. Still, the assumptions about the underlying causal network will often remain opaque, and thus, the conditions under which a coefficient can (or cannot) be interpreted as a causal effect remain unclear.

This opaqueness enables undesirable flexibility (e.g., Eisenberg, 1984; Smaldino, 2017), which discourages cumulative research. If researchers do not clearly specify the causal
effect they think they have identified, a study’s findings are hardly falsifiable. Imagine, for example, Researcher A publishes a nonexperimental study on subjective well-being and relationship satisfaction in which she concludes that a person’s low subjective well-being causes relationship dissatisfaction in the romantic partner. Researcher B might read the paper and disagree with the conclusion because Researcher B thinks the health of the person confounds the relationship between subjective well-being and the partner’s relationship satisfaction. Researcher B might then write a comment and criticize Researcher A’s study for not assessing and controlling for health, or Researcher B could conduct a new study to investigate whether the relationship still holds when controlling for health. On the other hand, if Researcher A had not explicitly claimed that the effect of subjective well-being on the partner’s relationship satisfaction was causal, Researcher B would have had a hard time pinning down what exactly to say about Researcher A’s study because “Your study did not correctly answer the question it did not explicitly try to answer” is not a compelling criticism. If confronted with criticism, Researcher A could retreat to the position that the finding was descriptive to begin with, even if this particular reading of her study is probably less interesting. Being unclear about the purpose of a study opens the door to such motte-and-bailey strategies in which researchers profit from the more interesting but difficult to defend causal interpretation of their effect (the bailey), but once challenged, they retreat to the almost trivial yet difficult to attack descriptive finding (the motte).

No single study can test all assumptions and rule out all potential alternative causal explanations. A variety of study designs, data sources, and methods are needed to attain confidence in estimates of causal effects (e.g., Briley, Livengood, & Derringer, 2018; Hernán, 2018b; Lawlor, Tilling, & Davey Smith, 2016). Such a cumulative endeavor needs to explicitly consider the assumptions that are involved, or else research may simply go around in circles, or end up in a futile back and forth when nobody notices that their discrepant
conclusions hinge on certain assumptions about which one could argue in a more fertile manner.

**Disconnect Between Original Findings and their Subsequent Interpretation**

The taboo against explicit causal reasoning and language has furthermore led to a disconnect between the original nonexperimental findings and their subsequent interpretation. Even if authors refrain from making causal interpretations in their original study, subsequent theoretical papers, reviews, or Introduction/Discussion sections will refer to the very same findings in a way that only makes sense if they were meant to be read as causal effects. The citing authors likely have no intention to mislead readers—they might simply have not considered the design of the respective study in great detail.

For example, on the empirical basis of longitudinal research that did not explicitly estimate causal effects, the neo-socioanalytic theory has proposed that investments in age-graded social roles drive (i.e., cause) personality trait change (e.g., entering the workforce after education leads to increases in conscientiousness; e.g., Roberts & Wood, 2006). Theories are usually causal in nature because cause-effect relationships permeate the way we think and make sense of the world (e.g., Kant, 1781/2002; Waldmann et al., 2006). Hence, if the empirical researchers in a field do not tackle causal questions explicitly and instead try to constrain themselves to descriptive or predictive statements and research questions, then a disconnect between empirical findings and theory is almost inevitable.

A similar disconnect can arise when nonexperimental studies are cited to make certain arguments in literature reviews and Introduction sections. For example, two recent reviews argued that intervention studies on how to change personality traits are vital and needed because personality traits predict important life outcomes in the domains of education, work, relationships, health, and well-being (Bleidorn et al., 2019; Roberts et al., 2017). The implicit assumption must be that personality traits cause the life outcomes; otherwise, changing the personality traits through interventions will not change the respective outcomes. It is possible
that personality is indeed the cause; however, most previous empirical studies on the topic did not explicitly investigate these causal effects.

Engaging in “stealth causal inference” from a distance (i.e., assuming causal relationships on the basis of descriptive or predictive findings reported elsewhere) may be convenient for nonexperimental fields because it means that authors do not have to defend explicit causal claims, yet everybody gets to enjoy explanatory accounts and the impression of a deep understanding of the subject matter. However, the disconnect between original findings and subsequent causal interpretations renders arguments and theories—even those that seem to be supported by an impressive number of empirical studies—speculative, which limits their usefulness for researchers and policy makers alike.

For researchers, speculative arguments and theories are not very helpful for designing causally informative studies because, even though speculative theories might stimulate new research ideas, they do not provide reliable information about which variables to assess and control for. Furthermore, the speculative nature of theories means that derived hypotheses have a lower prior probability of being true than the hypotheses derived from less speculative theories (i.e., theories with firmly established relationships and laws such as natural selection in Darwin’s theory of evolution). Their lower prior probability in turn results in more nonsignificant findings and more false positive findings (Diekmann, 2011; Fiedler, 2017; Ioannidis, 2005). Hence, theories without firmly established relationships and laws are not particularly useful, for example, for tackling the replication crisis in psychology (Fiedler, 2017; see also Muthukrishna & Henrich, 2019).

For policy makers, theories are useful if they contain firmly established causal relationships because only then are policy makers able to design interventions that

---

2 If the prior probability for each of 10,000 hypotheses H1 is .04, then there are 320 true positives and 480 false positives, assuming a statistical power of .80 and an alpha level of .05. If the prior probability is .20, then there are 1,600 true positives and 400 false positives. If the prior probability is .40, then there are 3,200 true positives and 300 false positives (Diekmann, 2011; Fiedler, 2017).
successfully tackle pressing issues in the world. Although predictive findings might help to identify at risk groups that might want to be targeted by interventions (e.g., adolescents with learning disabilities or self-control issues), predictive findings do not inform policy makers how they can intervene. We can thus understand one reason for the lamented lack of interventions and policies targeting personality traits (e.g., Bleidorn et al., 2019)—unless we establish that personality traits are indeed meaningful causes, why would one want to target them?

**Recommendations for Integrating Causality in a More Productive Manner**

**Steps of Causal Inference in Nonexperimental Studies**

How can we do better? Nonexperimental researchers should openly admit when their goal is causal inference—and then ensure that their study pursues this goal in a rigorous and transparent manner. The following four steps of causal inference might help them do so.

In Step 1, researchers should articulate a clear causal question and state the precise definition of the causal effect of interest. Translating the causal question into a hypothetical experiment and counterfactual thinking can help researchers do so because the counterfactual question “What would happen to an individual if one changed the treatment?” lies at the heart of causal inference (e.g., Foster, 2010b; Hernán, 2018a; Morgan & Winship, 2015). That is, the causal effect of interest is the difference between the outcome that did occur and the outcome that would have occurred had the individual not experienced the treatment (i.e., an event, condition, or characteristic). Thinking about how things are for an individual and how things would be different if the individual had not experienced the treatment can be formally expressed using the potential outcomes framework (for an accessible introduction, see Foster, 2010b; see also e.g., Holland, 1986; Rubin, 2005).

In Step 2, researchers might want to think carefully about how other variables relate to the putative causal variable (i.e., the treatment) and outcome variable to identify potential
confounders, colliders, mediators, and instrumental variables (see Figure 1). The assumptions about this underlying causal web can be expressed in a directed acyclic graph (DAG; e.g., Pearl, 2009). A DAG connects variables with arrows representing causal relationships. Importantly, the DAG should contain all relevant variables, not only the ones that are available, observable, or measurable. The DAG helps researchers align the study design and data analysis to the actual aim of the study (for accessible introductions to DAGs, see, e.g., Foster, 2010b; Rohrer, 2018). As a side note, whereas counterfactual thinking and DAGs may be new tools for many psychologists, they are in line with Campbell’s tradition of identifying plausible threats to internal validity (i.e., causal inference) and then including study design features and statistical adjustments that can potentially rule out those specific threats (e.g., Campbell, 1988; Matthay & Glymour, 2020; West & Thoemmes, 2010).

Step 3 involves establishing an identification strategy and estimating the causal effect. That is, given the assumptions from the previous steps, researchers derive a way to estimate the causal effect without bias from the data at hand. For example, this could involve a multiple regression model if all relevant confounding variables are available in the data, or it could involve the use of instrumental variable estimation (see Figure 1) if unobserved confounding is assumed (for introductions to and discussions of various identification strategies, see Foster, 2010b; MacKinnon & Pirlott, 2015; Mõttus & Kandler, 2018; Pingault et al., 2018; Rutter, 2007). Further inspiration for methods that can be used to investigate causal relationships on the basis of nonexperimental data can be found in fields such as economics, political science, or sociology. Parts of economics, political science, and sociology have embraced the challenge of causal inference on the basis of nonexperimental evidence, for example, through the use of instrumental variables estimation (see Figure 1),

---

3 A collider is a variable that is causally influenced by two variables, one that is (or is associated with) the treatment and another that is (or is associated with) the outcome. Adjusting for a collider can introduce a spurious (i.e., noncausal) association between the treatment and the outcome (e.g., Elwert & Winship, 2014; Rohrer, 2018).
regression-discontinuity designs, or fixed-effects models (e.g., Allison, 2009; Angrist & Pischke, 2008; Gangl, 2010; Morgan & Winship, 2015). All of these approaches have their own pitfalls, but psychologists are lucky that they can learn from critical discussions that have already transpired in other fields of research. Once the identification strategy is in place, it can be used to estimate the causal effect.

In the last step, Step 4, researchers test their identification strategy against violations of assumptions to see how much the effect estimate would change if certain assumptions were violated. For example, if the assumption is that all confounders have been observed, a researcher might want to compute what would happen if unobservable variables were to confound the effect (for more information on sensitivity analysis, see, e.g., Frank, Maroulis, Duong, & Kelcey, 2013; Greenland, 1996; Rosenbaum, 2005; Rosenberg, Xu, & Frank, 2019; VanderWeele & Ding, 2017). The last step should also involve a discussion of potential alternative explanations for the observed effect. This discussion, along with future directions for research, might be provided in the Discussion section.

In Figures 2, 3, S1, and S2, we briefly illustrate these four steps of causal inference with research questions from the four articles of Table 1. Please note that a detailed description and exemplification of all steps is beyond the scope of the current article (for more details on steps of causal inference, see Foster, 2010b).

**Further Recommendations**

Whereas the details of every particular attempt of causal inference will necessarily vary, we advise psychologists to be explicit about the whole process. Researchers should state that they are trying to estimate a causal effect, and they should be clear about the assumptions underlying their analyses. Being open about causality invites more critical reflection about the underlying assumptions, which may also open the door for more refined and productive rebuttals as points of disagreement can be pinpointed. To cite Charles Darwin (1981/1871), "False facts are highly injurious to the progress of science, for they often endure long; but
false views, if supported by some evidence, do little harm, for everyone takes a salutary pleasure in proving their falseness; and when this is done, one path towards error is closed and the road to truth is often at the same time opened” (p. 385).

Likewise, we advise researchers to make explicit rather than implicit causal inference statements in the arguments and theories they present in their Introduction sections, Discussions, reviews, and theoretical papers. This does not mean that they ought to make bold causal claims when there is substantial uncertainty. Instead, they should simply be more transparent about when an argument or theory depends on the existence of a particular causal effect (rather than just a correlation), and they should discuss the extent to which previous studies have provided compelling evidence for it. To do so, it might be helpful to state whether a causal effect in a theory or argument rests on previous experimental or nonexperimental evidence.

Finally, we suggest that the field as a whole should try to shift its norms toward a more productive engagement with causal inference on the basis of nonexperimental data. Statistics and methods teachers could dedicate some more time to the topic—it may be time spent well, as a clearer framework for causal inference makes it easier to talk about a broad range of topics, such as missing data problems (Thoemmes & Mohan, 2015) and threats to validity, which affect most types of research (Matthay & Glymour, 2020). Editors and reviewers may also encourage a shift in thinking. By no means should they let their guard down and allow researchers to confuse correlation with causation. However, instead of simply policing language or requesting boilerplate statements about limitations, they might ask hard questions—about the actual goal of the study (e.g., asking for clarification about why mere prediction would be interesting, or highlighting discrepancies between supposedly noncausal questions and the discussed implications), about the authors’ understanding of the underlying causal web (e.g., requesting that the authors provide a DAG to justify their choice of covariates), or about more specific recommendations for future studies (e.g., if an
experimental clarification is suggested, there should be some discussion about what a feasible experiment could look like). In some cases, authors may actually feel confident enough to make a causal claim—if it is accompanied by a transparent discussion of the underlying assumptions, then readers are given the information they need to form their own opinions.

**Conclusion**

Causal inference on the basis of observational data is very hard. However, this is not a good reason to render explicit causal inference taboo. Similar to when sex or drugs are made taboo, making explicit causal inference taboo does not stop people from doing it; they just do it in a less transparent, regulated, sophisticated, and informed way. Thus, we think it is about time that psychologists begin to talk openly about causality in nonexperimental research.
References


https://doi.org/10.4135/9781412993869


Kant, I. (2002). *Kritik der reinen Vernunft [Critique of pure reason]* (16th ed.). Frankfurt am Main, Germany: Suhrkamp. (Original work published 1781)


Table 1

<table>
<thead>
<tr>
<th>Place</th>
<th>Content of Text Passage</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example 1 – Moffitt et al. (2011)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>A gradient of childhood self-control predicts health, wealth, and public safety</td>
<td>The title indicates that the research question is not causal but predictive in nature (i.e., the study intends to predict health, wealth, and public safety). That said, when prediction is the goal, then the focus is usually not on individual predictors (e.g., self-control). Moffitt et al. first talk about self-control intervention programs to boost health and wealth and reduce crime. Given that intervention programs that increase self-control would only be effective (“reap benefits”) if self-control not only predicts but also causes these outcomes, the research question seems to be causal. The question about whether self-control is “important for the health, wealth, and public safety of the population” is vague. The subsequent sentence about prediction and the title of the article suggest that the study intends to investigate whether self-control is an important variable when one intends to predict health, wealth, and public safety. The terms “active ingredient” and “affecting” suggest that what is needed is causal knowledge. The term “causal” is absent. Plus, whereas they suggest that policy makers need causal knowledge, they do not explicitly state whether the goal of the study is to actually provide such knowledge.</td>
</tr>
<tr>
<td>Abstract</td>
<td>Policy-makers are considering large-scale programs aimed at self-control to improve citizens’ health and wealth and reduce crime. Yet, is self-control important for the health, wealth, and public safety of the population? Following a cohort of 1,000 children from birth to the age of 32 y, we show that childhood self-control predicts physical health, substance dependence, personal finances, and criminal offending outcomes, following a gradient of self-control.</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>Policy-making requires evidence that isolates self-control as the active ingredient affecting health, wealth, and crime, as opposed to other influences on children’s futures, such as their intelligence or social class origins. Dunedin study data allowed the requisite statistical controls for IQ and social class.</td>
<td>The statement that self-control is a clear target for intervention policy suggests that a “causal effect” is intended.</td>
</tr>
<tr>
<td>Comment section (i.e., Discussion section)</td>
<td>[…] it was possible to disentangle the effects of children’s self-control from effects of variation in the children’s intelligence, social class, and home lives of their families, thereby singling out self-control as a clear target for intervention policy.</td>
<td></td>
</tr>
</tbody>
</table>
**Current Research section**

Approximately 2,000 German students were tracked over 4 years from high school to university or to vocational training or work. […] First, the experience of life events over the 4 years of the study should be predicted by standing on personality traits at Time 1 (T1). […] Second, in line with the results reported by Vaidya et al. (2002), we hypothesized that experiencing more positive events would be associated with increases in extraversion, whereas experiencing negative events would be associated with increases in neuroticism.

The research goals seem to be descriptive (e.g., “students were tracked”; “experiencing more positive events would be associated with increases in extraversion”) and predictive (“be predicted by standing on personality traits at Time 1”).

---

**Results section**

[Heading:] Life Paths and Personality Traits: Selection and Socialization

In this section (and other parts of the manuscript), the authors talk about socialization effects, which implies that the authors intend to investigate the causal effects of life paths (studying at a university vs. vocational track) and life events on the Big Five personality traits. Only if the life events cause personality changes will they provide insights into why personality traits change. Hence, the interpretation of the findings in this passage is not in line with purely descriptive or predictive research goals.

---

**Discussion**

What was most compelling about our study of life events and their relation to personality development was how they provided insights going beyond any current theoretical ideas on why personality traits change in young adulthood.

---

**Example 3 – Grosz et al. (2019)**

Current Research section

First, mean-level changes in narcissistic admiration and Mach during early adulthood were examined in both cohorts (TOSCA-2006 and TOSCA-2002). Second, […] we investigated associations between studying economics and changes in narcissistic admiration and Mach. Third, […] we investigated associations between life events and changes in narcissistic admiration and Mach.

The first part of the study (mean-level changes) is descriptive. The second and third parts of the study also appear to be descriptive because the authors talk about investigating “associations.”

Method section

We ran the studying economics and life events analyses with and without control variables. We included the control variables for two reasons. First, we included them to prevent spurious associations. For example, the initial level of self-esteem might be a confounder […]

The inclusion of control variables and the mentioning of “spurious associations” and “confounders” would make more sense if the goal was to estimate a causal effect than if the goal was to simply describe the associations.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Example 4 – Study 1 from Cheng, Tracy, Foulsham, Kingstone, &amp; Henrich (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results section</td>
<td>[Heading:] Experiences Related to Changes in Narcissistic Admiration and Machiavellianism During Early Adulthood (Socialization Effects)</td>
<td>The goal is to investigate whether dominance and prestige are avenues to social rank and influence (i.e., whether dominance and prestige have a causal effect on social rank). This passage suggests that the study investigated the causal effect (“impact”) of the two strategies on social rank.</td>
</tr>
<tr>
<td>Limitations section</td>
<td>Finally, although we used the term socialization effect in this study in accordance with previous research on personality development, our data and analyses did not allow us to make causal claims.</td>
<td>The phrase “rank-attainment strategies” might imply that the study investigates whether Dominance and Prestige have a causal effect on rank-attainment.</td>
</tr>
<tr>
<td>Example 4 – Study 1</td>
<td>Two Ways to the Top: Evidence That Dominance and Prestige Are Distinct Yet Viable Avenues to Social Rank and Influence</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>In 2 studies, we investigated the impact of 2 fundamental strategies—Dominance (the use of force and intimidation to induce fear) and Prestige (the sharing of expertise or know-how to gain respect)—on the attainment of social rank, […]</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>In Study 1 we examined whether Dominance and Prestige spontaneously emerge and coexist as viable rank-attainment strategies within the same social groups, by asking previously unacquainted individuals to complete a collaborative task and allowing social hierarchies to naturally emerge.</td>
<td></td>
</tr>
<tr>
<td>Current Research section</td>
<td>One limitation of the present research is our reliance on a correlational approach, which prevents us from directly addressing questions of causality—whether Dominance and Prestige are causal antecedents to social rank.</td>
<td>Here, the authors follow the standard practice in nonexperimental psychology to avoid addressing causal research questions straightforwardly.</td>
</tr>
</tbody>
</table>
What is instrumental variable estimation?

Instrumental variable estimation is a method to estimate the causal effect of the treatment $X$ on the outcome $Y$ with the help of an instrumental variable $Z$. An instrumental variable $Z$ is a variable that is associated with the treatment, and only due to its association with the treatment, it is associated with the outcome. More specifically, an instrumental variable should fulfill the following four assumptions:

(a) The relevance assumption: The instrument $Z$ and treatment $X$ are associated either because $Z$ has a causal effect on $X$ (left panel), or because $X$ and $Z$ share a common cause $U^*$ (right panel).

(b) The exclusion restriction: $Z$ affects the outcome $Y$ only through $X$.

(c) The exchangeability assumption (also called independence assumption): $Z$ does not share common causes with $Y$ (other than $U^*$).

(d) The monotonicity assumption: $Z$ cannot increase $X$ for some individuals and decrease it for others (e.g., Bollen, 2012; Labrecque & Swanson, 2018; Lousdal, 2018).

The assumptions can only partially be tested empirically and require theoretical justification (Labrecque & Swanson, 2018). If an instrument that meets these assumptions can be identified, the causal average effect of $X$ on $Y$ can be estimated even in the presence of unmeasured confounding $U$. A variable that does not fulfill assumptions (b) and (c) can be transformed into a variable that fulfills these assumptions by adjusting for confounding variables.

For a continuous treatment variable, the instrumental variable estimand is the ratio: \[
\frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}
\]

Different types of instruments have been proposed: researcher induced randomization (e.g., a randomized anti-smoking intervention where the treatment is smoking), natural randomization processes (e.g., Mendelian randomization, where alleles are allocated at random in offspring), and natural variation (e.g., preference for treatment based on availability of facility or physician; e.g., Bollen, 2012; Lousdal, 2018).

\textit{Figure 1.} Basic information about instrumental variable estimation.
Step 1 (Basic definitions). Childhood self-control is the treatment variable that causes the outcome adulthood physical health. Self-control is defined as the ability to delay gratification, control impulses, and modulate emotional expression. Physical health is defined as cardiovascular, inflammatory, respiratory, dental, and sexual health (Moffitt et al., 2011).

Step 2 (Causal network). The DAG below illustrates the causal relationships we assumed on the basis of previous research. For example, serotonin levels in the central nervous system are believed to have a genetic basis, to be alterable by life circumstances, to affect conscientiousness (i.e., which is often seen as synonymous with self-control), and to help regulate the core bodily functions (appetite and sleep) that are necessary for good health (e.g., Carver, Johnson, Joormann, Kim, & Nam, 2011; Caspi, Hariri, Holmes, Uher, & Moffitt, 2010; Friedman et al., 2014). Importantly, genes, childhood socioeconomic status (SES1), and childhood serotonin are confounders because they have independent causal paths to youth self-control (SC1) and adulthood physical health.

Step 3 (Identification strategy). An approach that could be applied to block the confounding paths involving genes, SES1, and childhood serotonin and thus to get an unbiased estimate might be to run a study with monozygotic twins who are raised in the same family (i.e., pairs of individuals with identical genes and socioeconomic background). We would estimate the causal effect by regressing the intra-twin difference in adulthood health on the intra-twin difference in childhood self-control and the intra-twin difference in childhood serotonin. Differencing blocks the paths via the nodes genes and SES1 under the assumption that genes and SES1 influence both individuals of a twin pair in the same way (e.g., Allison, 2009; Campbell & Kenny, 1999; Kim & Steiner, 2019). Controlling for the intra-twin difference in childhood serotonin would neutralize the confounding effect of childhood serotonin.

Step 4 (Probing assumptions and alternative explanations). The validity of the identification strategy depends on whether all confounding twin-varying variables were included in the DAG and properly adjusted for in the analysis. For example, the intra-twin differences in childhood self-control might have been caused by intra-twin differences in adverse childhood experiences that caused also intra-twin differences in childhood.
intelligence, and childhood intelligence (not childhood self-control) might have been the actual cause of adulthood health. Thus, as a robustness check, we would add the intra-twin difference in childhood intelligence as control variable in the regression.

*Figure 2.* Brief illustration of the steps of causal inference for Example 1 (Moffitt et al., 2011).
**Step 1 (Basic definitions).** Studying at a university (as opposed to vocational training or work) at around age 20 to 25 is the treatment variable that causes the outcome conscientiousness at around age 25. Conscientiousness is defined as a personality trait characterized by the propensity to follow socially prescribed norms for impulse control, to be goal directed, to plan, and to be able to delay gratification (Roberts, Jackson, Fayard, Edmonds, & Meints, 2009).

**Step 2 (Causal network).** The DAG below illustrates the causal relationships we assumed on the basis of previous research. For example, the study by Spiess and Wrohlich (2010) suggests that the distance to the nearest university decreases the probability of enrolling in higher education.

![DAG Illustration](image)

*Note.* The numbers behind the variables indicate the time period. Consc = conscientiousness; SES = socioeconomic status; GPA = grade point average (high school); Snares = harmful lifestyles (e.g., started smoking, unplanned pregnancy).

**Step 3 (Identification strategy).** We propose that distance to the nearest university be used as an instrumental variable (for details on instrumental variable estimation, see Figure 1). That is, we would utilize the fact that the distance to the nearest university causally affects studying at a university, but that there are hardly any other paths from distance to nearest university to conscientiousness. The alternative paths all go through regional characteristics of the place of origin: Places far from any university might be poorer or more rural. The poorness or ruralness of a place might cause both whether there is a university and the likelihood of studying via the socioeconomic status of the family of the participant (e.g., parents’ education). Thus, we would control for the regional characteristics (indicators of urbaneness and prosperity of the place of origin) in the instrumental variable estimation.

**Step 4 (Probing assumptions and alternative explanations).** One assumption is that all paths from distance to the nearest university to adulthood conscientiousness go through studying at university and regional characteristics. To probe this assumption, we would regress youth conscientiousness on distance to the nearest university and regional characteristics. If distance to the nearest university would be incrementally associated with youth conscientiousness, this would suggest that there are paths from distance to nearest university to adulthood conscientiousness that are not mediated by regional characteristics or studying at university, which would question the validity of the instrumental variable (i.e., distance to the nearest university).

*Figure 3.* Brief illustration of the steps of causal inference for Example 2 (Lüdtke et al., 2011).
**Step 1 (Basic definitions).** Studying economics at a university is the treatment variable that causes Machiavellianism (Mach) in later life. Studying economics is defined as studying business, macroeconomics, or another economics major as opposed to choosing a noneconomics major (e.g., engineering, sociology). Mach is defined as a personality trait characterized by cynical world view, pragmatic morality, cold rationality, deviousness, and a belief in the effectiveness of (one’s own) manipulative tactics (e.g., Christie & Geis, 1970; Jones & Paulhus, 2009).

**Step 2 (Causal network).** The DAG below illustrates the causal relationships we assumed mainly on the basis of previous research.

**Note.** The numbers behind the variables indicate the time period. SES = socioeconomic status; GPA = grade point average.

**Step 3 (Identification strategy).** To attain a suitable instrumental variable for estimating the causal effect of studying economics on Mach, we would randomly assign participants to one of two conditions. In the experimental condition, participants would receive an advertisement about studying economics at a university in their final year of high school. In the control condition, participants would receive an advertisement about studying noneconomic majors. The random assignment to condition ensures that the advertisement is not confounded with other causes of Mach (see DAG). Furthermore, we assume that the advertisement affects the level of Mach after studying at a university only via studying economics. Admittedly, the advertisement might also have an effect on the level of Mach immediately after the advertisement, but this effect should be short-lived (i.e., temporary state changes in Mach) and should thus not influence the level of Mach after studying at a university.

**Step 4 (Probing assumptions and alternative explanations).** A potential limitation of the design is that those who are swayed by the advertisement might also be those who are swayed by the study major to change their personality (i.e., either due to the content of the courses or due to influences from peers). Thus, the estimation of the causal effect might only be valid for the subpopulations that responded to the advertisement (i.e., local average treatment effect) but not for other subpopulations of young adults. Another potential limitation might be that the advertisement has a very weak effect on choice of study major (“weak instrument”), which could lead to biased and inconsistent estimates (e.g., Bound, Jaeger, & Baker, 1995; see also the relevance assumption in Figure 1). Furthermore, if the advertisement backfires among some participants (e.g., if the provided information makes...
studying economics less attractive for some participants), the monotonicity assumption underlying the instrumental variable is violated (Figure 1).

*Figure SI*. Brief illustration of the steps of causal inference for Example 3 (Grosz et al., 2019).
**Step 1 (Basic definitions).** Dominance is the treatment variable that causes the outcome social rank. Dominance is defined as the use of force and intimidation to induce fear in others (Cheng et al., 2013). Social rank is defined as the amount of social influence on others and attention from others in face-to-face groups.

**Step 2 (Causal network).** The DAG below illustrates the causal relationships we assumed mainly on the basis of previous research. For example, previous research suggests that the height and muscularity of an individual might cause both dominance and social rank in face-to-face groups (e.g., Melamed, 1992; Blaker & Van Vugt, 2014).

**Step 3 (Identification strategy).** We propose that the paths that might confound the estimation of the causal effect be blocked. We would block the path from sex to social rank by using only same-sex groups as Cheng et al. (2013) did. We would block the other paths by including muscularity, height, voice pitch, facial width-to-height ratio, physical attractiveness, prestige, and age as control variables in a multilevel model—we would use multilevel modeling because each individual is nested in a face-to-face group (Cheng et al.).

**Step 4 (Probing assumptions and alternative explanations).** Because the identification strategy is to condition on observable variables, it is particularly important that all relevant variables are in the DAG. If a relevant variable is not part of the DAG, this might invalidate the identification strategy. For example, if physical health would have a causal effect on both dominance and social rank, then physical health might confound the association between dominance and social rank. Another assumption is that all variables in the model were assessed without measurement error. This assumption might not be valid for some variables (e.g., dominance, social rank, muscularity). We would check the robustness of our estimation by running the multilevel analysis in a structural equation model and correcting for measurement error by using latent variable modeling (e.g., Westfall & Yarkoni, 2016).

*Figure S2.* Brief illustration of the steps of causal inference for Example 4 (Cheng et al., 2013).