Crud (Re)Defined

Amy Orben1,2 and Daniël Lakens3

1MRC Cognition and Brain Sciences Unit, University of Cambridge; 2Emmanuel College, University of Cambridge; and 3Human-Technology Interaction Group, Eindhoven University of Technology

Abstract
The idea that in behavioral research everything correlates with everything else was a niche area of the scientific literature for more than half a century. With the increasing availability of large data sets in psychology, the “crud” factor has, however, become more relevant than ever before. When referenced in empirical work, it is often used by researchers to discount minute—but statistically significant—effects that are deemed too small to be considered meaningful. This review tracks the history of the crud factor and examines how its use in the psychological- and behavioral-science literature has developed to this day. We highlight a common and deep-seated lack of understanding about what the crud factor is and discuss whether it can be proven to exist or estimated and how it should be interpreted. This lack of understanding makes the crud factor a convenient tool for psychologists to use to disregard unwanted results, even though the presence of a crud factor should be a large inconvenience for the discipline. To inspire a concerted effort to take the crud factor more seriously, we clarify the definitions of important concepts, highlight current pitfalls, and pose questions that need to be addressed to ultimately improve understanding of the crud factor. Such work will be necessary to develop the crud factor into a useful concept encouraging improved psychological research.

Keywords
crud factor, effect sizes, null-hypothesis significance testing, review, open data, open materials

Meehl’s concept of the “ crud factor”—the idea that everything correlates with everything else in psychological science—is currently experiencing a renaissance. Discussions about this 50-year-old concept were largely confined to statistical commentaries and debates about null-hypothesis significance testing for decades. In recent years, the increased focus on research practices and large-scale data in the psychological sciences has, however, gradually pushed the crud factor out of the shadows of scientific attention. In turn, its use in the psychological research literature has increased.

Regrettably, the renewed popularity is at least partly due to the flexibility with which the crud factor is used in the scientific literature. Although its idea is compelling, the crud factor lacks a clear definition, and as a result, many reports fail to accurately represent its inherent complexities. Consequently, when doing so is convenient, researchers use the crud factor to dismiss statistically significant findings, either in their own work (e.g., Zondag, 2013) or in the work of others (e.g., Elson et al., 2019). In its original form, however, the crud factor is highly inconvenient for any psychological scientist. It puts into question the use of common statistical significance tests that rely on the absolute null hypothesis (Meehl, 1990a, 1990b). When a concept that has inconvenient consequences for a field’s dominant approach to statistical inference has evolved into a convenient tool for researchers to discount unwanted findings, something has gone awry.

This review tracks the development of the crud factor in the scientific literature, with a focus on its definition (or more specifically, the lack thereof) and use in practice. We formalize Meehl’s original definition of the crud factor, discuss challenges in providing evidence of its existence, and present a review that maps crud’s

Corresponding Author:
Amy Orben, Emmanuel College, University of Cambridge, Cambridge CB2 3AP, United Kingdom
E-mail: aco35@cam.ac.uk
presence in the scientific literature to date. We thereby uncover the diverse conceptualizations and uses of the crud factor and the problems that emerge from the current lack of scientific consensus. Finally, we raise specific questions that, when answered, would improve how the crud factor is utilized, explored, and applied in future psychological studies.

What Exactly Is the Crud Factor?

Published work discussing the crud factor has played an important role in increasing awareness about the idea that in psychology everything correlates with everything else. It is difficult to give credit for this idea to any single researcher. Meehl, who first mentioned the term in the literature (Meehl, 1984), credited Lykken (Lykken, 1967, 1984, 1986, 1990a, 1990b). In turn, Lykken, who initially referred to the concept as “ambient noise” (Lykken, 1968), credited Meehl’s unpublished work and previous researchers (e.g., Bakan, 1966; Nunnally, 1960). The popular assumption that the concept of the crud factor was devised by Meehl is therefore an over-simplification. However, he played a major role in its popularization, as one of its most vocal advocates.

The crud factor lacks a clear and enforceable definition despite its popularity. In 1966 and 1968, Lykken defined its precursor—ambient noise—as the average shared variance between unrelated variables. He highlighted several factors that might contribute to this variance, including (a) the frequent correlation of positive psychological and physical variables, (b) the fact that experimenters can unintentionally bias the results of their studies, (c) common method variance (the underlying correlation between a participant’s data points if the same method was used to elicit them), and (d) the fact that state variables, such as fatigue or anxiety, can affect all measures in an experimental session. Lykken’s definition of ambient noise therefore encompassed both correlational and experimental work.

Lykken’s definition remains controversial and seems to contradict Fisher’s (1925) notion that randomization provides a formal justification for the usefulness of null-hypothesis significance tests and related statistical methods. If, as Lykken’s definition indicates, even randomized interventions influence other variables and thereby set in motion complex and nontheorized chains of relationships, which in turn can change the outcome measurement, the absence of systematic differences between experimental conditions is no longer guaranteed.

For example, consider an experiment in which the presentation of social and nonsocial stimuli is used as a manipulation of perspective taking and the question of interest is whether perspective taking influences subsequent performance in constructing a persuasive essay. Lykken might have argued that even if the perspective taking of participants is not affected by whether the stimuli are social or nonsocial, there is an almost infinite range of potential confounds that are unrelated to this aspect of the stimuli but lead to small differences in the content of the pictures or their visual properties. These differences can affect essay-writing performance through, for example, influences on affect, arousal level, or motivation.

Meehl (1967) initially supported Lykken’s idea that even experimental randomization cannot guarantee the truth of the absolute null hypothesis, given that “everything in the brain is connected with everything else,” and therefore that it is “highly unlikely that any psychologically discriminable stimulation which we apply to an experimental subject would exert literally zero effect upon any aspect of his performance” (p. 109). However, Meehl avoided such statements in his later publications, writing instead that the crud factor is an issue only for correlational work and for experimental work that relies specifically on interpreting interactions between the intervention and nonrandomized variables (Meehl, 1990a, 1990b, 1997). He noted disagreements between himself and Lykken (Meehl, 1990a, 1990b, 1997). Specifically, he argued that “everything is correlated with everything else, more or less,” but distanced himself from Lykken’s belief that causally “everything influences everything” (Meehl, 1990a, p. 123). It must be noted here that for experiments not to be affected the crud factor, one would need to implement a perfect randomization and intervention procedure that perfectly manipulates only the variable of interest. If this is not the case, reasoning based on experimental results can also be subject to the influence of multivariate complex causal pathways that were not theorized by the experimenters.

Meehl also ceased to consider the influence of experimenter bias, common method variance, and state variables as part of the crud factor, distancing his definition from Lykken’s (1968) definition of ambient noise. For example, he left out common measurement variance from his definition of the crud factor (Meehl, 1990b), but acknowledged that the underlying quality of measurement partly determines the magnitude of a data set’s crud factor (Lykken, 1991; Meehl, 1990a). Meehl (1990b) instead stressed that the crud factor consists of “real differences, real correlations, real trends and patterns for which there is . . . some true but complicated multivariate causal theory” (p. 208). Factors that could account for such an all-encompassing causal structure include, for example, genes, child-rearing, environmental factors, intelligence, and social class (Meehl, 1990a).

In this framework, crud-factor relations are real and replicable correlations that are themselves subject to sampling error and that often reflect true, but complex, multivariate and nontheorized causal relationships.
Before Crud

Authors often have overlooked the fact that the idea of a null hypothesis of exactly zero is always false has been discussed in the scientific literature for more than half a century. In a commentary about chi-square tests published in 1938, Berkson conjectured that if a normal curve is fitted to data “representing any real observations whatever of quantities in the physical world, then if the number of observations is extremely large—for instance, on the order of 200,000—the chi-square \( P \) will be small beyond any usual limit of significance” (p. 526). He concluded that a chi-square test is “no test at all!” (p. 527). Nunnally (1960) extended this position to between-variable correlations, arguing that “in the real world the null hypothesis is almost never true” because “just as nature abhors a vacuum, it probably abhors zero correlations between variables” (pp. 642, 643).

The term common sense is a hallmark of this early literature about the crud factor. Nunnally (1960), for example, noted that his position was “supported both by common sense and by practical experience,” but acknowledged that this is ultimately “a personal point of view [that] cannot be proven directly” (p. 642). Meehl relied on very similar argumentation years later: “There is nothing mysterious about the fact that in psychology and sociology everything correlates with everything” (Meehl, 1990b, p. 204) as this is obvious “from the armchair on common-sense considerations” (Meehl, 1990a, p. 123). Possibly taking it to the extreme, Bakan (1966) wrote that the idea of the null hypothesis always being false is so commonsense that to put it in writing “is, as it were, to assume the role of the child who pointed out that the emperor was really outfitted only in his underwear” (p. 423).

Failed Attempts to Confirm the Crud Factor’s Existence

This reliance on “common sense” rests, in part, on the inability to ultimately confirm or disconfirm the crud factor’s existence. To falsify the notion of the crud factor, for example, one would need to empirically falsify the prediction that the absolute null hypothesis is always false (Berger & Delampady, 1987). To do this, one would need to find correlations in nonrandomized data sets that are truly zero; yet it might well be practically impossible to ever show a correlation that is close enough to zero to satisfy a true skeptic. Furthermore, it is impossible to collect enough data to conclusively determine whether the effect of interest is exactly zero or just extremely small.

It is also not possible to generalize that the crud factor exists from previous work. Though early supporters of the idea completed studies showing that all correlations between all variables in large data sets reach statistical significance (Bakan, 1966), such findings cannot be extrapolated to all psychological data sets. Although these tests are therefore far from conclusive, we summarize this work to demonstrate that the amount of evidence for omnipresent correlations between all variables in a data set varies across research domains.

Personality psychology is among the areas with the most evidence for correlations between all variables in a data set. In an unpublished study in 1966, Meehl and Lykken examined data from the University of Minnesota Student Counselling Bureau’s Statewide Testing Program, which gave a questionnaire to 57,000 high-school seniors. For 15 selected variables, the researchers calculated 105 different correlations and chi-square values. All reached statistical significance, and 96% were significant to the 10^-6 level (Meehl, 1990b). A study of 25,000 Mayo Clinic participants’ data found that 507 out of 550 items showed a sex difference (Swenson, Pearson, & Osborne, 1973).

In a replication of such work, 135 variables were examined in data from 2,058 grade-school students, and, on average, a given variable correlated significantly with about 41% of the other variables in the data set (Standing, Sproule, & Khouzam, 1991). The lower percentage of significance in this study could be due to the data set’s smaller sample size, but it raises questions about possible inconsistencies. More recently, Waller (2004) specified random directional sex-difference hypotheses by running simulations on data from 81,485 individuals who completed the 567-item second revised version of the Minnesota Multiphasic Personality Inventory (MMPI). Forty-six percent of the simulated hypotheses reached statistical significance. If this finding extends to psychological data sets more generally, it would back up Meehl’s (1967) idea that directional hypotheses should be supported 50% of the time when tested in large-enough samples.

These studies have, however, mainly been limited to showing that if statistical power is sufficiently high, most correlations in the data sets tested become statistically significant, but the studies have not provided more generalizable evidence. Efforts to confirm the crud factor’s existence are further complicated by the absence of theoretical and conceptual work on what the crud factor represents, which makes it difficult to know how to parse the true crud factor from the general background of common method variance and experimental biases. The crud factor is therefore a concept that cannot be proven to exist using current methodologies.

The lack of conceptual debate about and empirical research on the crud factor has been noted by critics who disagree with how some scientists treat the crud
factor as an “axiom that needs no testing” (Mulaik, Raju, & Harshman, 1997, p. 80). They worry that such axiomatic treatment of the crud factor could “introduce an internal contradiction into statistical reasoning” that hinders researchers’ ability to think and reason about the world (Mulaik et al., p. 80). To prevent this, such critics stress that scientists should consider the crud factor as an “empirical generalisation” that needs to be tested (Mulaik et al., 1997, p. 80). We believe it might be possible to parse the epistemological, and nonprovable, concept of the crud factor from its estimation-based and possibly quantifiable counterpart, which we call the crud estimate, thereby creating a framework that can be tested and used in future research. To that end, we clarify definitions of the two terms in the next section.

**Definitions**

**Crud factor**

The crud factor still lacks a concrete and commonly accepted definition: Meehl never formally defined the crud factor, and few scientists have attempted to develop a standardized definition since (e.g., Levine, Weber, Hullett, Park, & Lindsey, 2008). Researchers can therefore select their preferred definition of the crud factor or—as most do—decide not to include a conceptual definition at all, relying instead on quoting Meehl (1990b): “Everything correlates with everything to some extent” (p. 207). We define the crud factor as follows:

Crud factor: The epistemological concept that, in correlational research, all variables are connected through causal structures, which result in real nonzero correlations between all variables in any given data set.

With this definition, we aim to summarize the ideas Meehl described in his original work and put them in concrete terms, even though Meehl himself never synthesized these ideas into a single definition (Meehl, 1990a, 1990b). The definition encompasses data sets collected in the behavioral and psychological sciences and focuses on causal structures that include both direct and indirect pathways linking two variables.

**Crud estimate**

Even though there is little empirical proof for the generalizability of the claim that in any given data set all variables have nonzero correlations, there have already been endeavors to quantify the size of the crud factor in different research areas. Some researchers have used such size estimates to discount results in their own or others’ studies that fall beyond this cutoff, often without recognizing the lack of evidence for the cutoff they use. Such occurrences are not surprising as the term *crud factor* misleadingly alludes to the existence of a numerical factor. However, estimation processes could be a valuable avenue for developing a testable and quantifiable aspect of the crud factor. To dissociate the epistemological concept of the crud factor from such a numerical entity, we provide a separate definition of the crud estimate:

Crud estimate: An estimate of the average correlation between variables in a multivariate data set that are linked by nontheorized and true causal structures.

**Quantifying the Crud Estimate**

Pinpointing an estimate of the average correlation between variables linked by nontheorized causal structures is no easy task. First, there is no consensus about whether it is even possible to determine a crud estimate, given that there is little theory to help differentiate crud from other factors, such as common method variance. Second, to provide a crud estimate, one would need to distinguish theoretically meaningful correlations from correlations driven by causal structures that are not of interest theoretically, even though, from the outside, these correlations look identical (Webster & Starbuck, 1988). Third, this estimate will be different for different research domains, as noted by Meehl (1990a): “Doubtless the average correlation of any randomly picked pair of variables in social science depends on the domain and also on the instruments employed” (p. 124).

Meehl and Lykken provided educated guesses of the crud estimate in their early work; Lykken (1968) thought that “it is not unreasonable to suppose” that ambient noise has a correlation of .2 (p. 153), whereas Meehl (1986) initially believed that the crud factor “can hardly be supposed to be less than . . . r = .25 in the soft areas of psychology” (p. 327). In 1990, Meehl examined pairwise correlations between MMPI scales and found an average correlation of about .3, which he then assumed to be a viable crud estimate (Meehl, 1990a, 1990b). Such estimates will not be universal; they will depend on the area of psychological study and the nature and psychometric properties of the measurements taken. Other studies have, for example, shown lower average correlation values (e.g., $r = .07$; Standing et al., 1991). Furthermore, average correlations fail to provide ultimate answers: It is still unclear how much of an average correlation value is driven by the theorized as opposed to nontheorized correlations. Researchers referencing crud estimates in their work without considering the (lack of) evidence for these estimates are therefore
building their reasoning on exceptionally unstable foundations.

There is, however, a potential approach that could help researchers crudely quantify a crud estimate for the data sets they are working with: They could discount the correlations between variables in their data set that they deem theoretically meaningful and examine all remaining correlations. One way to estimate which relations in a data set are not theoretically meaningful could be to ask experts in the relevant field of research to indicate how a priori likely they believe each possible relation between variables in the data set to be. They could rate the extent to which they believe there is a causal and theoretically meaningful relationship between each pair of measurements in the data set. If they state that they do believe there could be a theoretically meaningful relationship between two variables, they could then be asked to explain why. If multiple experts go through this process, one could meta-analyze the observed correlations between all pairs of variables (assuming these are measured with very high accuracy) for which the experts collectively failed to acknowledge any theoretically meaningful relationship (e.g., the 10% of correlations with the lowest ratings). It must be noted, however, that such a method comes with considerable measurement challenges, and substantial theory development might be required in some research areas before experts can come to reasonable agreement. Furthermore, this method will still have difficulties distinguishing between common method variance and true crud. Despite these practical challenges, we believe that, in principle, a quantification of the crud estimate might be possible.

Disclosures

Our codebook, coding sheet, and code to create the figures in this article can be found on the Open Science Framework (OSF; https://osf.io/uebp8/).

Review of Past Literature

To understand the role the crud factor and crud estimate play in psychological science, we needed to unravel how their use has developed over time. We therefore applied a systematic search and coding procedure to review all the psychological literature since the late 1900s in which the crud factor was used for research or commentary purposes. Our procedure was closely based on the PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009), but did not adhere to the parts of the guidelines concerning quantifying a relationship through a preregistered meta-analytic method. This is because our review was conducted not to quantify relationships but as a scoping exercise to improve current understanding of how the crud factor has been used in past literature. Our systematic review was last updated on the October 3, 2019.

We used Google Scholar to search for the terms ['Crud Factor' OR 'Crud factor' OR 'crud factor'] and ['crud' AND 'Meehl'] to ensure that we examined all articles, books, book chapters, working papers, reports, preprints, and theses that included the concept anywhere in their text (not just in their abstract, keywords, or title). We also searched for ['ambient noise' AND 'Lykken'] to catch those works that possibly miscited the crud factor, given that ambient noise and crud factor are often referenced in conjunction. We supplemented this search with a search on dimensions.ai, which hosts some more diverse research outputs, with help of an institutional librarian, but did not contact authors to identify additional studies. We initially identified 288 scholarly works (see Fig. 1).

The initial exclusions were done solely by the first author. She first removed all duplicates and those works not in English. She also excluded those records that did not reference the crud factor as it is known in psychological and behavioral sciences (e.g., aerospace engineering also has a “crud factor”). She thereby identified 211 scholarly works that were screened fully for their eligibility. At that point, she excluded 18 additional works that only mentioned “ambient noise” or that were written by Meehl while initially developing the concept of the crud factor. Another 4 were not accessible and were therefore excluded. Our final sample, after search and exclusion, consisted of 189 works, spanning 1990 to 2019. The sections of these works that referred to the crud factor were extracted and coded by the first author using the codebook available on OSF. In short, the coder examined each of these sections to code whether the definition of the crud factor used by the author referred to it as representing “not real” or “trivial” relationships, whether the definition included concepts of common method variance, and whether the only attempted definition was to quote Meehl as saying that “everything correlates with everything.” The coder also noted whether a crud estimate was used in the study. No methods were used to identify risk of bias in the studies. The second author coded 37 (~20%) of the works and agreed with the first author on 91% of the coding instances (i.e., 168 out of 185).

Uses of the crud factor

The sample of scholarly works shows that the popularity of the crud factor has increased in recent years, but is still not very high. The greatest number of citations occurred in 2018, when it was cited fewer than 20 times (Fig. 2).
During the review, it became evident that the largest proportions of works in the sample were in the areas of statistics and psychometrics, an indication of the term’s popularity in the methodological literature. Yet we saw a recent boost outside these areas. The crud factor has therefore crossed the methodological boundary and is now increasingly applied in wider areas of original research and commentary.

We proceeded to examine how many scholarly works in the sample used the term crud factor in a way that is not consistent with Meehl’s original definition, formalized in this article. Although inconsistent use of the term is not inherently a negative research attribute, many works in the sample used a definition of the crud factor that was more convenient for the argumentation of those specific works. In 10.1% of the works, the crud factor was equated with measurement errors, such as common method variance. Although common method variance was a component of Lykken’s concept of ambient noise, Meehl’s original crud factor encompassed only real correlations, and therefore any interpretation of crud as common method variance is inconsistent with Meehl’s definition.

The crud factor was defined as being “trivial” or “not real” in 12.7% of the works in our review. However, the crud factor does not represent “spurious correlations” (Alexander, 2018, p. 55) that are “statistical artifacts” (Benge, Onwuegbuzie, & Robbins, 2012, p. 92); it is...
wrong to say that the crud factor does “not represent authentic or unique associations” (Adams, 2012, p. 142).

Apart from the coding, we noted that other works labeled the crud factor as being a result of experimental error or miscited the concept. Furthermore, there were unique inconsistencies with Meehl’s definition. For example, the crud factor was referred to as a problem only for nonaggregated data (Ibarra, 2016), a result of scientists’ preference for significant results (Delgado-Romero & Howard, 2005), a reason why psychologists can still interpret correlations at small sample sizes (Marchese, 1992), and something that can be avoided by using power analysis (Kouvelis & Kouvelis, 1998) or a Bonferroni correction (Sotelo & Sangrador, 1999). More than a quarter of the scholarly works (25.4%) based their definition of the crud factor solely on Meehl’s popular quote that “everything correlates with everything else.” But there were also many works that did not include any sort of definition or defined the crud factor so vaguely that it was difficult to judge how the authors intended to define the term.

**Uses of the crud estimate**

In addition to examining the use of the crud factor in the literature, we investigated the occurrence of the crud estimate. Only 33 scholarly works in our review included a numerical crud estimate, and many of those referenced the crud estimates suggested by Lykken or Meehl (e.g., Meehl’s 1990 estimate of \( r = .3 \)). Most works in our review failed to caution readers that these estimates were educated guesses, lacking any solid empirical data because there is still little “accurate knowledge about the size of the crud factor in any given research domain” (Cohen, 1994, p. 1000).

Quite a few of the works also incorrectly referenced Meehl (1990) to back up other crud estimates. For example, one commentary stated that “standardized paths should be at least 0.20 and ideally above 0.30 in order to be considered meaningful [in structural equation modeling]” (Chin, 1998, p. xiii). Still others suggested completely different guidelines, without citing any supporting evidence. For example, one article claimed that researchers should treat “significant correlations between .15 and .25 [in psychometrics] . . . as marginally acceptable but only if they are theoretically predicted and are therefore less likely to be part of the inexplicable crud” (McKelvie, 1994, p. 1224).

Overall, most of the works in our sample lacked clear explanations of why a certain crud estimate was used. Most problematic was the tendency for researchers to pick a numerical estimate that was convenient for them. For example, one author discounted an unwanted correlation of .18 as being within the crud-factor range, but interpreted another correlation of .24 as statistically meaningful (Zondag, 2013).

**Approaches to Crud in the Age of Big-Data Psychology**

Psychological researchers are currently accessing ever-increasing amounts of data, and are therefore obtaining an ever-increasing ability to detect small effects. There is therefore an omnipresent risk that the crud factor will be routinely labeled as statistically and theoretically significant (Orben & Przybyliski, 2019). Evidence originating from very convoluted multivariate causal structures could then be taken as evidence for hypothesized theories, even when this evidence cannot be attributed to the proposed causal structure (Meehl, 1990b).

In the current scientific climate, higher statistical power is often equated with better-quality evidence, but the possible existence of a crud factor should reinforce researchers’ vigilance about stating that a theory or hypothesis is supported because of a single successful statistical significance test. It is important to investigate the crud factor and crud estimate in more detail, by developing new methods for understanding or quantifying them. Until this need is addressed, there will still be outstanding key questions about how psychological researchers should adjust their use of both the crud factor and the crud estimate. These questions concern the clarity of definitions, the choice of the crud estimate, the benefit of complex models and precise predictions, and the need to avoid using the crud factor as a tool of convenience:

- **Clarity:** How can researchers ensure that the crud factor is used more consistently? It is currently possible for every researcher to use idiosyncratic working definitions of the crud factor and crud estimate, so that no two researchers use the same definitions. We therefore recommend that authors always detail their preferred definitions in writing and refrain from solely citing a source. Furthermore, the fact that the crud factor represents real but nontheorized relationships, and what that implies for reported results, should be described and evaluated.

- **Choice:** How can researchers choose a crud estimate that is relevant for their research area? We have described a possible approach to quantifying the crud estimate, but we must note here that there might need to be considerable theory development in the psychological and behavioral sciences before this approach can be deemed feasible. A less sophisticated way of substantiating a crud estimate has been applied in the organizational-psychology
literature (Webster & Starbuck, 1988) and in a study about the design of massive open online courses (MOOCs; van der Sluis, van der Zee, & Ginn, 2017). In this work, the distributions of background correlations were provided in histograms. Although such a histogram does not perfectly portray a crud estimate, it could be used to examine whether an observed effect size is more likely to reflect a real effect or crud-like relations. A likelihood ratio could quantify whether an observed correlation is more likely to have originated from a distribution centered around the hypothesized magnitude of the effect size or from crud. Yet more work is clearly needed to substantiate the usefulness of such an approach, especially given that the distributions do not successfully separate the crud factor from other relations, such as common method variance. This work would also have to recognize and account for the fact that interpreting a crud-type relationship as reflecting a theorized one might be more harmful in certain fields of research than in others; for example, it might be more harmful in research directly informing practice or policy than in basic exploratory research.

• Severity: How can researchers safeguard themselves against the crud factor while more targeted procedures are developed? One approach might be to construct and test theories that make more falsifiable predictions, either because their predictions are more complex or because they specify the magnitude, not just the direction, of the expected effect. The risk of the crud factor influencing the evaluation of a theory can be substantially reduced if the theory provides a point estimate or multiple directional predictions (Meehl, 1967, 1990a, 1990b). In essence, researchers should consider the severity of their test (Mayo, 2018) and make predictions that can be proven wrong. Prediction of any nonzero effect has almost no risk of being falsified if the crud factor plays a role, so researchers need to carefully consider, and preferably specify in advance, what would falsify their prediction.

• Convenience: How can researchers move away from referencing the crud factor only when it is convenient for a scientific argument? Fully addressing this question will be possible only after a better understanding of the crud factor is achieved. One possible way to encourage more general adoption of the crud factor into quantitative research could be for journals to adopt approaches similar to the one put forward by McKelvie in 1994. He suggested that if an effect size falls below a specified crud estimate, it should be considered a worthwhile scientific result only if it was theoretically predicted; if it was unpredicted, it should be treated as inconsequential. We want to emphasize that any implementation of such a method must take into account that some worthwhile scientific effects might be smaller than the crud estimate of the data set. If clear predictions of effect sizes are made in research, taking this into account would not be difficult. One could, for example, determine whether an observed effect is closer to the hypothesized effect size or to the range of relations that are likely to be driven by crud. Such an approach could ensure that the crud factor is not used only by those researchers who benefit from it.

Conclusions

The crud factor is increasingly being used as a tool to guide the interpretation and dismissal of minute effects. To ensure that it does not turn into an “ever-convenient crap factor” (Mossbridge & Radin, 2018, p. 113), which can be defined and estimated by researchers to discount those effects they do not like in their own or others’ work, psychological researchers need to collaborate to achieve scientific consensus about the crud factor’s existence, definition, estimation, and use. By providing an initial definition and introducing suggestions for next steps, we have underlined the importance of critically evaluating this increasingly popular concept in the psychological sciences—however convenient it might be.

Transparency

Action Editor: Alexa Tullett
Editor: Daniel J. Simons
Author Contributions
A. Orben and D. Lakens conceptualized the idea for this study and performed the review. A. Orben drafted the manuscript, and D. Lakens gave important feedback throughout.

Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding
The authors thank the European Association of Social Psychology and the British Psychological Society for funding A. Orben in conducting this research during a visit to D. Lakens. D. Lakens was funded by VIDI Grant 452-17-013 from the Netherlands Organisation for Scientific Research.

Open Practices
Open Data: https://osf.io/uebp8/
Open Materials: https://osf.io/uebp8/
Preregistered: no
All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/uebp8/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/2515245920917961. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

**ORCID iDs**

Amy Orben https://orcid.org/0000-0002-2937-4183

Daniël Lakens https://orcid.org/0000-0002-0247-239X

**Acknowledgments**

We want to thank the members of Daniël Lakens’s lab and its associates at other universities for giving crucial feedback during the writing process.

**References**

Adams, T. G., Jr. (2012). The unique roles of affect and impulsivity in the prediction of skin picking severity. *Journal of Obsessive-Compulsive and Related Disorders, 1*, 138–143.

Alexander, L. A. (2018). Good question? How subtle changes in item wording alters correlations between cognitive symptom measures of anxiety and depression (Publication No. 10686753) [Doctoral dissertation, George Mason University]. ProQuest Dissertations and Theses Global.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66*, 423–437. doi:10.1037/h0020412

Benge, C. L., Onwuegbuzie, A. J., & Robbins, M. E. (2012). A model for presenting threats to legitimation at the planning and interpretation phases in the quantitative, qualitative, and mixed research components of a dissertation. *International Journal of Education, 4*(4), 65–124.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science, 2*, 317–352.

Berko, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association, 33*, 526–536.

Chin, W. W. (1998). Commentary: Issues and opinion on structural equation modeling. *MIS Quarterly, 22*(1), vii–xvi.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*, 997–1003.

Delgado-Romero, E. A., & Howard, G. S. (2005). Finding and correcting flawed research literatures. *The Humanistic Psychologist, 33*, 293–303.

Eison, M., Ferguson, C. J., Gregerson, M., Hogg, J. L., Ivory, J., Klisanin, D., . . ., Wilson, J. (2019). Do policy statements on media effects faithfully represent the science? *Advances in Methods and Practices in Psychological Science, 2*, 12–25.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Ibarra, L. (2016). Statistics vs scientific explanation. *Mankind Quarterly, 56*, 374–383.

Kouveliotus, A., & Kouveliotus, I. (1998). Relations among measures of job satisfaction, role conflict, and role ambiguity for a sample of Greek teachers. *Psychological Reports, 82*, 151–163.

Levine, T. R., Weber, R., Hultlett, C., Park, H. S., & Lindsey, L. L. M. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research, 34*, 171–187.

Lykken, D. T. (1966). *Statistical significance in psychiatric research*. Retrieved from https://conservancy.umn.edu/bitstream/handle/11299/151660/PR-66-9.pdf?sequence=1

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.

Lykken, D. T. (1991). What’s wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest* (pp. 3–39). Minneapolis: University of Minnesota Press.

Marchese, M. C. (1992). *An empirical investigation into the construct redundancy of job evaluation and job re-design* (Doctoral dissertation). Retrieved from https://lib.dr.iastate.edu/rtd/9849/

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge, England: Cambridge University Press.

Mckelvie, S. J. (1994). Guidelines for judging psychometric properties of imagery questionnaires as research instruments: A quantitative proposal. *Perceptual and Motor Skills, 79*, 1219–1231.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103–115.

Meehl, P. E. (1984). Foreword. In D. Faust (Ed.), *The limits of scientific reasoning* (pp. 11–25). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). What social scientists don’t understand. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 315–338). Chicago, IL: University of Chicago Press.

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108–141.

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*, 195–244.

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393–425). Mahwah, NJ: Erlbaum.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine, 6*(7), Article e1000097. doi: 10.1371/journal.pmed.1000097

Mossbridge, J. A., & Radin, D. (2018). Plausibility, statistical interpretations, physical mechanisms and a new outlook: Response to commentaries on a precognition review. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 110–116.
Mulaik, S. A., Raju, N. S., & Harshman, R. A. (1997). There is a time and a place for significance testing. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 65–116). Mahwah, NJ: Erlbaum.

Nunnally, J. (1960). The place of statistics in psychology. *Educational and Psychological Measurement, 20*, 641–650.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour, 3*, 173–182. doi:10.1038/s41562-018-0506-1

Sotelo, M. J., & Sangrador, J. L. (1999). Correlations of self-ratings of attitude towards violent groups with measures of personality, self-esteem, and moral reasoning. *Psychological Reports, 84*, 558–560.

Standing, L., Sproule, R., & Khouzam, N. (1991). Empirical statistics: IV. Illustrating Meehl’s sixth law of soft psychology: Everything correlates with everything. *Psychological Reports, 69*, 123–126.

Swenson, W. M., Pearson, J. S., & Osborne, D. (1973). *An MMPI source book: Basic item, scale, and pattern data on 50,000 medical patients.* Minneapolis: University of Minnesota Press.

van der Sluis, F., van der Zee, T., & Ginn, J. (2017). Learning about learning at scale: Methodological challenges and recommendations. In *L@S ’17: Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (pp. 131–140). Retrieved from https://doi.org/10.1145/3051457.3051461

Waller, N. G. (2004). The fallacy of the null hypothesis in soft psychology. *Applied and Preventive Psychology, 11*, 83–86.

Webster, J., & Starbuck, W. H. (1988). Theory building in industrial and organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 93–138). Chichester, England: Wiley.

Zondag, H. J. (2013). Narcissism and boredom revisited: An exploration of correlates of overt and covert narcissism among Dutch university students. *Psychological Reports, 112*, 563–576.