Where Science Meets Discourse: What a Flawed Commentary of Three Papers Can Teach Us About Research on Well-Being in the Digital Age

Andrew K. Przybylski*[1] and Matti Vuorre[2]

[1]Oxford Internet Institute, University of Oxford
[2]Department of Social Psychology, Tilburg University

Abstract

Research examining technology and psychological well-being has become increasingly important for health policy, international regulation, and behavioral science. A notable consequence of this increased attention has been an increasingly commentary-driven public discourse where influence and research contribution and careful analysis are not always proportionally aligned. While commentary can be useful, it can also introduce misunderstandings into the public, research, and policy ecosystems if it is not grounded in rigorous argumentation and empirical observation. Criticism lacking these qualities can nonetheless present valuable opportunities to address misunderstandings and improve science communication. In this paper we examine one such commentary on three of our papers. We address the four issues raised and clarify how each either misunderstands or misrepresents our work, and then translate these errors into broader lessons for those interested in understanding, conducting, and communicating behavioral research in the digital age.

Background

Digital and internet-based technologies are changing the world. Good, careful science investigating the technical, social and psychological dynamics of those technologies is important, as the young people of today are growing up in a world heavily mediated through the internet, smartphones, and social media. Unfortunately, too much of the discussion involves overconfident, unevidenced claims about technology's associations with well-being and mental health originating from Netflix, Substack, and popular press books. These claims are presented as empirically sound, but do not have the rigor required of scientific inquiry (Odgers, 2024). A recent essay (Sigaud et al., 2025) provides a clear example: The commentary's conclusions are suggestive of a scientific critique of three academic papers (Vuorre & Przybylski, 2023a, 2023b, 2024; henceforth VP2023a, VP2023b, and VP2024) but the argument and evidence provided lack elements of genuine scientific inquiry.

The commentators raised four issues:

1. "The 'Blender' Problem": A suggestion that we "blended" data inappropriately, leading to unclear concepts or methods.
2. "Facebook's Influence on Research Design": An allegation that industry involvement affected our study's design or outcome.
3. "Problematic Use of Two Datasets": A claim that two of the datasets we studied, or how we did so, were unsuitable for the studies.
4. "Inability to Infer Causation": A note that our studies were correlational and exaggerated conclusions.

In this article we address these points, and by doing so we show how the critiques miss the mark and highlight how careful science communication and academic rigor can improve

---

* Corresponding author: andy.przybylski@oii.ox.ac.uk. We thank David Ellis, Karen Mansfield, Margarita Panayiotou, and Doug Parry for feedback on an earlier draft.

public understanding in an area which is increasingly dominated by polemics and powerful interests.

1. How We Described Research ("The 'Blender' Problem")

We start with what the commentators call the "blender" problem. They suggest we mixed different types of data and demographics and thereby muddied our concepts and methods. This criticism reveals a misunderstanding of our work and, possibly, modern behavioral science methods in general. Specifically, the commentators assert:

> "[Vuorre and Przybylski] combine data spanning different years, ages, countries, and technologies together and estimate an average correlation between digital technology and psychological well-being or mental health—an approach that we will refer to as 'blending.'" (Sigaud et al., 2025, p.4)

We discuss these four claims in turn below.

Blending years

> "Because Vuorre and Przybylski report results for all years *combined*, they may be diluting a stronger signal of harm in the later years." (p.6)

We did not "report results for all years combined" in VP2024. Instead, we estimated separate associations for each year with "random intercepts and slopes over years and the country-by-year interaction because any associations might be heterogeneous over time (Vuorre et al., 2021)." (VP2024, p.3 and Equation 1). As shown in Figure 2 in VP2024, which includes all years' coefficients, results did not meaningfully differ by year.

We did not emphasize potential variation in associations over time in VP2023a and VP2023b because our earlier report—ignored by the commentators—found few changes in associations between technology use and mental health (Vuorre et al., 2021). Because we had this empirically informed prior expectation, we did not interact associations with time but instead attempted to adjust for it in VP2023a and VP2023b.

Blending ages

> "Vuorre and Przybylski's papers provide few insights into [*children and teens*, especially girls and those in early puberty] because the studies blend the youngest participants into pools mostly composed of older participants" (p.4)

While (we hope) informative about young individuals, our three works did not exclusively address technology use's relations with teenage girls' well-being, for reasons evident to readers of our papers: VP2023a used a 15-34-year-old category due to limitations in the data as discussed in the manuscript, VP2023b and VP2024 reported on all ages in the datasets.

The commentators note that we found "suggestive evidence of beneficial effects" (p.5; we did not suggest our estimated parameters are *effects*) among 10-14 and 15-19 year old females in the Global Burden of Disease (GBD) dataset, but dismiss this finding as problematic due to limitations in the GBD data (which we already described in VP2023b, p.9). (They also ignore our corresponding results from the Gallup World Poll [GWP] dataset.) One overarching goal in our reports was comprehensiveness—something the commentators seemed to eschew in favor of a simple and therefore possibly misleadingly vague narrative by selective reporting in their commentary.

In VP2024, for a recent and more inclusive definition of adolescence (Sawyer et al., 2018) and computational efficiency, we used an age category of 15-24 because under 15-year-olds are not present in the data. We communicated this clearly and remain uncertain as to what new information regarding age "blending" the commentary aims to provide.

Blending countries

> "Blending together all countries may dilute a stronger signal of harm…" (p.6)

The equations in each of our three manuscripts describe how we estimated country-specific associations. Figure 2 in each manuscript shows the country-specific parameters. The only "blending" in our studies is a desirable feature of multilevel models (Efron & Morris, 1977).

Curiously, a guest post on the commentary authors' blog* criticizes our work for having figures with "several hundred uninterpretable squiggly lines" showing countries separately, so we are uncertain as to whether the commentators are against: i. presenting results separately per country, ii. presenting results in aggregate across countries, or iii. empirical reporting that doesn't agree with their favored narrative.

Blending technologies

We were puzzled by the commentators' suggestion that we "blended" different technologies and thereby obscured potential relations between psychological health and social media use. In the titles, abstracts, and bodies of our manuscripts we specifically talk about "Facebook adoption" (geography-specific daily/monthly active Facebook users; VM2023a), "adoption of Internet technologies" (geography-specific internet users and mobile broadband subscriptions; VM2023b), and "having (mobile) internet access or actively using the internet" (individuals' survey responses; VM2024). Only a misreading of our work can lead to the view that we blended technologies (we analyzed indicators separately) in a way that misunderestimated something we didn't study (teenage girls' social media use, presumably).

Ironically, the commentators' made-up term "blender problem" introduces the very kind of unclear terminology we tried to avoid. In behavioral science, we have established terms such as "random slopes", "secondary data analysis", and so on. They are not self-explanatory, but they aim to be precise—and precision is paramount to understanding how technology use relates to human well-being.

2. How We Analyzed a Dataset ("Facebook's Influence on Research Design")

The commentators' second issue is that Facebook provided part of the data we analyzed in VP2023a. They claim this influenced our findings. They highlight that our analysis did not include engagement data from every country in the world and that we did not focus on differences between different "generations". The commentators assert:

> "First, *it was Facebook — not the authors* — that chose the 72 countries to include in the sample. Why those 72 countries, and not other countries or all countries? Details about how these 72 countries were selected are needed to know how representative these findings might be." (Sigaud et al., 2025, p.8)

It is true that we did not have data on all countries. We did not make a universal claim or attempt to answer the questions the commentators are posing. The commentators are not making a novel point, but one that we already provided:

> "We also did not make attempts at finding a socially or geographically representative sample of nations to study, but rather used data from countries that Facebook determined to have the most accurate data about adoption and demographics. It is therefore possible that these results would not generalize beyond the sample of 72 nations we studied." (Vuorre & Przybylski, 2023a, p.7)

---

* https://www.afterbabel.com/p/gallup-world-poll. That blog post also contains criticisms of the GWP dataset, which, according to Gallup researchers, are "based on analytical errors and poor reasoning" (https://www.afterbabel.com/p/a-debate-on-the-strengths-limitations).

In a similar vein, the commentators assert:

> "Facebook made the decision to define the youngest age group as 13-34 years old, allegedly to 'maximize the accuracy of the data.' But by dictating this wide age group, Facebook guaranteed that this would be a study primarily of *adults in the Millennial generation*, not of Gen Z teens. In 2008, there were essentially no members of Gen Z on Facebook. (Gen Z begins with those born in 1996). Even in the last year of the dataset, 2019, the wide age range means that everyone from age 24 to 34 was a Millennial adult. To address the psychological questions at issue in this debate it would have been far preferable to limit the dataset to teens (say, ages 13-18, or even 13-21), so that researchers could track how Facebook affected teens as the generations changed." (Sigaud et al., 2025, p.8)

This point has something to do with American conceptualizations of generations, another popular buzzword that is scientifically fraught.* We did not set out to study generations and were clear about how this methodological decision was taken:

> "Facebook calculates [Daily and Monthly Active Use] estimates separately for individuals aged 13–34 and 35+. User age is determined based on Facebook profile information, which can be inaccurate (e.g. young users reporting an older age). Accordingly, Facebook has trimmed 0.008% of total MAU to exclude accounts with unrealistic or non-reported ages. Facebook chose the 13–34 and 35+ age categories in order to maximize the accuracy of the data." (Vuorre & Przybylski, 2023a, p.2)

This is not to say that we believed that the data comprehensively covered how those of different ages may have been interacting with social media over time. Unlike the idea of generations, there is good scientific reason to think that different social media platforms' users' ages have shifted over time. As such, we underscored a related but substantive point:

> "We also highlight the fact that while Facebook adoption remains the overall dominant social media platform, our results do not necessarily generalize across different platforms. For instance, in the United States, 13- to 17-year-olds are more likely to use TikTok, Instagram and Snapchat than Facebook, so the user base of Facebook now consists of relatively more older individuals." (Vuorre & Przybylski, 2023a, p.7-8)

It is indeed possible that Facebook conspired to mislead us. They might have anticipated how we were going to measure health and well-being and hatched a plan to provide bespoke age cohort and country data. We could have been misled when we believed their claim that data quality was better for these countries and demographics. Remarkably, this scheme would then have yielded the same basic pattern of results we observed from other secondary data analyses that did not rely on industry data. We have no way of knowing if this was the case.

For those who had questions about the data that we could not answer, such as theories proposed by the commentators, we noted the following in the paper:

> "The Facebook adoption data were made available to us on Facebook's Open Research Tool platform. Other researchers can contact Facebook (ccobb@fb.com) to access the dataset." (Vuorre & Przybylski, 2023a, p.3)

3. How We Used Two Datasets ("Problematic Use of Two Datasets")

The commentators' third issue centers mainly on our use of the Gallup World Poll (GWP) to "detect mental health changes" (p.9). We did not use the GWP to address questions about mental health but nevertheless attempt to respond to the criticism here. The commentators assert:

---

* For example, see https://aeon.co/essays/generational-labels-are-lazy-useless-and-just-plain-wrong and https://www.washingtonpost.com/opinions/2021/07/07/generation-labels-mean-nothing-retire-them/.

> "…the GWP has limited sample sizes for specific demographic subgroups, averaging only about 40 respondents per country/year/sex/age group. Such small samples risk statistical noise swamping meaningful signals. Consequently, the GWP is poorly suited to detect mental health changes in demographic groups most affected by social media." (p.9)

If our goal was to use the GWP to "detect mental health changes" it would not be suitable. Thankfully, we described our research questions in each paper. These provide a helpful guide to what we were investigating. In "Estimating the association between Facebook adoption and well-being in 72 countries" our focus was on well-being and for this we used the GWP:

> "We joined these unique datasets to conduct a descriptive study to answer three basic yet important questions. First, to what extent is Facebook adoption associated with well-being? Second, do these associations differ by age or sex. And finally, how might these associations have differed between countries? In addition, we were interested in whether the intensity of use might make a difference, and therefore conducted our analyses separately for daily active users and monthly active users. As a supplementary analysis, we replicated and present these analyses on meta-analytic mental health outcomes in an appendix [using GBD data]. Due to the exploratory and descriptive nature of our study, we did not have *a priori* hypotheses about the directions or magnitudes of the potential associations." (VP2023a, p.2)

In "Global Well-Being and Mental Health in the Internet Age" we reported two studies. The first was focused on well-being and for this we used the GWP:

> "In the first [study using GWP data], we focus on three aspects of psychological well-being and contrast them with yearly per capita Internet users and mobile-broadband subscriptions across 168 countries and 16 years. In the second [using GBD data], we examine three mental-health outcomes across 202 countries and 19 years. Our aim is to better understand (a) how well-being and mental health have changed, on a global scale, during the past 2 decades of dramatic proliferation of Internet technologies and connectivity; (b) how per capita Internet users and mobile-broadband subscriptions predict country-level well-being and mental health within a given country and across countries; and (c) the extent to which associations between Internet-technology adoption and well-being and mental health differ across age and sex and if they are specific to previously suggested vulnerable populations, such as young women." (VP2023b, p.1-2)

In "A Multiverse Analysis of the Associations Between Internet Use and Well-Being", our goal was to

> "…estimate the extent to which internet access, mobile internet access, and active internet use predicts psychological well-being on a global level. […] We studied eight indicators of well-being: life satisfaction, the extent to which individuals reported experiencing daily negative and positive experiences; two indices of social well-being; physical well-being, community well-being, and experiences of purpose. […] Our current research questions, then, were as follows: (a) To what extent does well-being differ between individuals who report having access to, or using, (mobile) internet? and (b) How robust are these differences in well-being across different internet adoption predictors, well-being outcomes, subgroups, and model covariate specifications?" (VP2024, p.2)

The commentators make a similar set of claims about how we operationalized well-being:

> "Additionally, Vuorre and Przybylski's analyses relied primarily on three outcomes—positive experiences, negative experiences, and life satisfaction—that vary in their relevance to mental health. For instance, their "negative experiences" category includes a question about feeling sadness the day before (relevant to internalizing disorders like depression and anxiety), anger (less directly relevant), and physical pain (irrelevant).

> Combining these diverse states could obscure important signals specific to anxiety or depression." (p.9)

Our use of the GWP cannot have obscured anything having to do with anxiety, depression, or self-harm because we did not use it to measure those things.* Instead, we used as few as three (VP2023a) and as many as eight indices (VP2024) of well-being to measure well-being and used estimates of anxiety and depression from a different dataset (the Global Burden of Disease, or GBD) in separate studies. Moreover, the commentators are either misreading or misrepresenting our work when they claim:

> "Vuorre and Przybylski themselves acknowledge that these measures [GWP well-being indicators] have 'not [been] extensively validated' as mental health indicators." (p.9)

Although we did not describe these limitations as having to do with "mental health indicators" we did note something similar about measures of well-being:

> "Although not extensively validated, we believe these items and scales to be uniquely valuable for our goals due to their extensive scope across time, geography and demographics. […] In addition to inclusion in the data, we were motivated to use this [life satisfaction] scale due to its widespread use, and because of prior work establishing its reliability and validity." (VP2023a, p.3)

> "While these scales are not psychometrically validated, Gallup cites prominent scientists as having helped with their development (Gallup, 2022, p. 5). While this statement does not make up for these scales' lack of validation, we believe that the extensive scope of the data set, across both time and countries, makes them uniquely valuable objects of study." (VP2024, p.3)

Moreover, we clearly delineated between our studies on, and the datasets used to address, well-being and mental health. For example:

> "Nevertheless, this analysis [Study 1] was necessarily constrained to a limited range of available outcomes reflecting subjective well-being (e.g., Jebb et al., 2020). In Study 2, we extended our investigation to mental-health outcomes." (VP2023b, p.9)

The commentators provide two ancillary points about how global mental health estimates are used in one of the three papers. But these are not new points and are detailed in the study in question. Specifically, they note it is a limitation that:

> "The GBD relies extensively on statistical modeling and imputation – *not actual data collection* – to derive its estimates" (p.9)

This is not a novel observation, but clearly described in our paper and references therein (VP2023b, p.9):

> "The GBD collates heterogeneous data from all World Health Organization member states' censuses, household surveys, civil registration and vital statistics, disease registries, health-service-use statistics, disease notifications, and other sources. It then aggregates data from these sources with Bayesian metaregression to produce country-specific yearly prevalence estimates.
>
> The GBD 2019 prevalence-rate estimates are based on 19,773 data sources with varying coverage for individual countries; for details of the GBD 2019 methodology, see Vos et al. (2020) and especially Appendix 1 therein." (VP2023b, p.9)

Likewise, the commentators cite their own blog post that questions the accuracy of the GBD:

* To prevent misunderstandings, we note that, contrary to the commentators' suggestions, anger and pain are related to mental health problems such as depression (e.g. Judd et al., 2013; Korff & Simon, 1996).

> “Notably, the GBD does *not* track known trends for similar outcomes in more reliable datasets. For example, Rausch (2024) documented large discrepancies between the GBD data and trends observed in high-quality datasets in the U.S., U.K., Australia, and France.” (p.9)

This is not a novel observation about the GBD. We wrote:

> “We emphasize that the GBD estimates are not observed data and therefore are accurate only to the extent that the GBD’s data-collection methods and modeling strategies are valid. We compared the GBD estimates with the Centers for Disease Control and Prevention’s (2022) estimates of self-harm in the United States and found that they are likely to deviate in systematic ways from other authoritative information sources. We nevertheless argue that because the GBD provides the most comprehensive data set of global mental health, studying these estimates is informative but emphasize this caveat.” (VP2023b, p.9)

Measuring well-being, mental health, and technological engagement over time and broad demographic and geographic groups is difficult. It is made harder, not easier, by commentators conflating different constructs or by falsely accusing others of doing so.

4. Caution to Infer Causation (“Inability to Infer Causation”)

The final issue raised by the commentators touches on an important topic for the field at large—causal inference—but does not add anything that is not common knowledge already. Interestingly, by raising this issue, the commentators show a misunderstanding of the concept that is itself informative. As in earlier sections, this issue is inaccurately framed by a suggestion that our studies failed at a research objective they did not have:

> “All three studies document correlations which should not be interpreted as causal.” (p.10)

This is accurate, but redundant to readers of our papers. Each explicitly communicated the non-causal nature of our designs and findings multiple times:

> “…our descriptive analyses cannot and do not rule out the possibility of causal effects, either negative or positive, between social media use and well-being.” (VP2023a, p.7)

> “Note that the data and theory required to address this question at the causal level are absent. Consequently, our analyses cannot account for potential confounders in the associations linking mental health and well-being to Internet-technology adoption. Our descriptions, therefore, are suggestive but are not intended to provide evidence for or against causal relations.” (VM2023b, p.15)

> “Finally, we further highlight the tentative-at-most nature of our results with respect to causal effects of internet access and use on individuals’ well-being. Causal inference from observational data, such as that studied here, is notoriously difficult (Rohrer, 2018; VanderWeele et al., 2016). Critical theoretical assumptions must be made and properly applied in the statistical models in order to approach unbiased causal estimates, steps that we did not take in the current work. We nevertheless remain hopeful that the clarity with which we hoped to address this issue will provide a solid foundation for future work on internet technologies’ causal effects (Grosz et al., 2020).” (VP2024, p.7)

The commentators make a seemingly related statement, but one that is flawed in a different way that merits attention:

> “The authors adjust for some individual-level socioeconomic variables in V&P 2024, but V&P 2023a and V&P 2023b make no adjustments for time-varying country-specific characteristics like GDP, unemployment rates, schooling in the population, or median household income.” (p.10)

Indeed, we do report models that do and do not include variables to test the robustness of our analyses. Nevertheless, the statement that we adjusted in VP2024 but not in VP2023a&b is itself misleading. Because this statement omits the rationales present in each paper it inaccurately suggests our modeling strategies were either incomplete or haphazard attempts to address the same question. They were not. The purpose of VP2024 was to explore the influence of those adjustments, whereas the latter papers' intent was to examine the unadjusted associations. Again, there is nothing new in the commentary about this issue that is not clearly laid out in our three papers. Ours was incremental scientific research where we made many well documented analytic choices and provided rationales and context in the text of the papers. None of these align with what the commentators suggest.

This misrepresentation of our papers speaks to a more fundamental confusion the commentary has about how different kinds of scientific evidence are diagnostic (or not) to the study of digital technologies and well-being. For example:

> "The biggest threat to causal inference here is the possibility of confounding by variables that are left out of their models." (p.10).

Putting aside the fact that we clearly communicated our non-causal aims, the commentators are highlighting not one but two challenges to causal inference by making this assertion. First, it highlights how hard it is to distinguish between research that provides evidence for causal claims and research that doesn't:

> "…or observational studies leveraging quasi-experimental variations in social media exposure (see Braghieri et al., 2022)—has provided evidence of social media's negative effects on mental health." (p.10)

Because the commentators do not recognize consistency between our non-causal research questions and our appropriately applied statistical methodology, they incorrectly conclude we failed at causal inference. In contrast, because the commentators do not recognize inconsistency between the causal research questions and the misapplied applied statistical methodology they incorrectly conclude the study they cite provides evidence that social media has negative effects on mental health (Eckles, 2023). This lack of discernment is a repeating pattern.

Second, the commentators apparently do not appreciate that selective attention to omitted variable bias has implications for the narrative-driven verbal theories of social media effects promulgated elsewhere. Although our papers never set out to:

> "…contradict empirical claims that heavy social media use causes increases in anxiety and depression among adolescents, and especially among girls." (p.10)

We are aware that some of the commentators regularly make confident claims like this. Many scholars have reminded them that the cornerstone of good causal inference is articulating precise, testable models to account for the complexities of the developmental, environmental, social, and psychological factors that shape mental health (University of Virginia, 2024)[*]. Nonetheless, the omitted variable bias issue raised by the commentators is an important one. Fortunately, there is a relevant paper on this exact topic that criticizes the commentators' related work: "Social Media and Youth Mental Health: Simple Narratives Produce Biased Interpretations" (Sewall & Parry, 2024). It provides a direct and concise treatment of this problem and:

[*] Conversation with Candice Odgers & Jon Haidt: "Making Sense of the Research on Social Media & Youth Mental Health" (https://www.youtube.com/watch?v=Ewxe4pWOH-I).

> "…[demonstrates] how one can obtain biased effects showing that social media use negatively affects mental health when the data-generating process is misspecified…" (Sewall & Parry 2024, p. 508).

Sewall and Parry statistically investigate how omitted confounding variables can bias estimates of causal effects and conclude that:

> "The prevailing narrative that social media use is the principal cause of declining youth mental health oversimplifies a complex phenomenon. This perspective draws attention away from the broader set of factors contributing to youth mental health problems and suffers from an omitted variable bias that can lead to biased conclusions." (Sewall & Parry, 2024, p. 512)

We strongly endorse Sewall and Parry's stance that academic scholars "are not arguing that social media platforms be given free reign nor are we dismissing the genuine harms that can occur online" (p. 513). Given the legal, ethical, and scientific stakes, now is an ideal time to set aside simple narratives and quick fix solutions:

> "Moving forward, research and policy efforts should adopt a more nuanced approach that acknowledges the multifaceted nature of mental health and considers the broader context in which social media use occurs." (Sewall & Parry 2024, p. 513).

Closing

We are at a crucial moment in studying digital technology and its effects on societies and individuals. We firmly believe that precision matters when using terminology, conducting analyses, sourcing diverse data, and interpreting causal claims. The easy path forward is to adopt the appearance, but not the substance, of the scientific process when determining the nature and scope of technologies' harms or benefits. The harder one and we think more important path—the scientific way—is to do the work while acknowledging its limitations: design rigorous studies, formulate substantive critiques, collect data at a global scale, analyze it transparently and comprehensively, and report and interpret findings accurately, even when they do not align with popular or personally favored narratives.

Contrary to the commentators' claims, our papers did what we sought to: Advance understanding of technology's relationships with well-being and represent incremental steps in an ongoing scientific process. They demonstrate transparent and robust methodological approaches for studying psychological and technological phenomena. Future research must build on genuine scientific inquiry rather than retreat to emotionally charged polemics. The societal impact of technology requires rigorous, scalable research methods that keep human rights in perspective. The importance of these issues demands that public discourse be informed by evidence from rigorous scientific work instead of the boosterism, doomerism, big tech lobbying, speculation and anecdote rife in the field today. It remains an open question how scholars can most effectively communicate the value of scientific research grounded in rigorous designs that test specific claims, transparency, and accurate reporting. It is our hope this article presents a positive step in this direction.

References

Eckles, D. (2023, August 22). thefacebook and mental health trends: Harvard and Suffolk County Community College. *Statistical Modeling, Causal Inference, and Social Science*. https://statmodeling.stat.columbia.edu/2023/08/22/thefacebook-and-mental-health-trends-harvard-and-suffolk-community-college/

Efron, B., & Morris, C. (1977). Stein's Paradox in Statistics. *Scientific American*, *236*(5), 119–127.

Judd, L. L., Schettler, P. J., Coryell, W., Akiskal, H. S., & Fiedorowicz, J. G. (2013). Overt Irritability/Anger in Unipolar Major Depressive Episodes: Past and Current Characteristics and Implications for Long-term Course. *JAMA Psychiatry*, *70*(11), 1171–1180. https://doi.org/10.1001/jamapsychiatry.2013.1957

Korff, M. V., & Simon, G. (1996). The Relationship Between Pain and Depression. *The British Journal of Psychiatry*, *168*(S30), 101–108. https://doi.org/10.1192/S0007125000298474

Odgers, C. L. (2024). The great rewiring: Is social media really behind an epidemic of teenage mental illness? *Nature*, *628*(8006), 29–30. https://doi.org/10.1038/d41586-024-00902-2

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health*, *2*(3), 223–228. https://doi.org/10.1016/S2352-4642(18)30022-1

Sewall, C. J. R., & Parry, D. A. (2024). Social media and youth mental health: Simple narratives produce biased interpretations. *Journal of Psychopathology and Clinical Science*, *133*(7), 507–514. https://doi.org/10.1037/abn0000950

Sigaud, L., Rausch, Z., McClean, A., & Haidt, J. (2025). *How Three Studies by Vuorre and Przybylski May Have Obscured the Impact of Social Media on Youth Mental Health*. OSF. https://doi.org/10.31234/osf.io/5fpeu_v1

University of Virginia. (2024). Conversation with Candice Odgers & Jon Haidt: “Making Sense of the Research on Social Media & Youth Mental Health.” *TYDE*. https://tyde.virginia.edu/livestream-odgers-haidt/

Vuorre, M., Orben, A., & Przybylski, A. K. (2021). There Is No Evidence That Associations Between Adolescents’ Digital Technology Engagement and Mental Health Problems Have Increased. *Clinical Psychological Science*, 2167702621994549. https://doi.org/10.1177/2167702621994549

Vuorre, M., & Przybylski, A. K. (2023a). Estimating the association between Facebook adoption and well-being in 72 countries. *Royal Society Open Science*, *10*(8), 221451. https://doi.org/10.1098/rsos.221451

Vuorre, M., & Przybylski, A. K. (2023b). Global Well-Being and Mental Health in the Internet Age. *Clinical Psychological Science*, 21677026231207791. https://doi.org/10.1177/21677026231207791

Vuorre, M., & Przybylski, A. K. (2024). A Multiverse Analysis of the Associations Between Internet Use and Well-Being. *Technology, Mind, and Behavior*, *5*(2). https://doi.org/10.1037/tmb0000127

Transparency

AKP is supported is financially supported by the Huo Family Foundation, UK Research and Innovation, and the Economic and Social Research Council. In previous five years AKP's research was also funded by the University of Oxford's John Fell Fund. From 2022 to 2024 AKP contributed as a paid scientific advisor to the Sync Digital Wellbeing Program. AKP conducts all research in line with the University's academic integrity code of practice.

MV's research is or has in the past five years been financially supported by the University of Oxford's John Fell Fund, the Economic and Social Research Council, and Tilburg University. MV has provided unpaid consultations and served as a non-paid panel member for Meta and K-Games. MV conducts all research in line with the Netherlands and European Code of Conduct for Research Integrity.