



How should we investigate variation in the relation between social media and well-being?

Niklas Johannes¹, Philipp K. Masur², Matti Vuorre^{1,3}, and Andrew K. Przybylski¹

¹Oxford Internet Institute, University of Oxford

²Department of Communication Science, Vrije Universiteit Amsterdam

³Department of Social Psychology, Tilburg University

Most researchers studying the relation between social media use and well-being find small to no associations, yet policymakers and public stakeholders keep asking for more evidence. One way the field is reacting is by inspecting the variation around average relations—with the goal of describing individual social media users. Here, we argue that this approach produces findings that are not as informative as they could be. Our analysis begins by describing how the field got to this point. Then, we explain the problems with the current approach of studying variation and how it loses sight of one of the most important goals of a quantitative social science: generalizing from a sample to a population. We propose a principled approach to quantify, interpret, and explain variation in average relations by: (1) conducting model comparisons, (2) defining a region of practical equivalence and testing the theoretical distribution of relations against that region, (3) defining a smallest effect size of interest and comparing it against the theoretical distribution. We close with recommendations to either study moderators as systematic factors that explain variation or to commit to a person-specific approach and conduct $N = 1$ studies and qualitative research.

Keywords: Social media, well-being, effect heterogeneity

The study of social media use and well-being is at a critical junction. The evidence seems clear: The relationship between social media use and the average user's well-being is close to zero and below any threshold for harm that would require intervention (Appel et al., 2020; Bayer et al., 2020; Dickson et al., 2019; Dienlin & Johannes, 2020; Kross et al., 2021; Odgers & Jensen, 2020; Orben, 2020b; Valkenburg, 2022; Valkenburg et al., 2022; vanden Abeele, 2020). Yet that consensus has not reached the public, which continues to express concerns over the harmful effects of media use (Grimes et al., 2008; Orben, 2020a), prompting governments to keep asking for more evidence (Council on Communications and Media, 2016; Dickson et al., 2019), and, in some cases, act without it (BBC, 2021). The disconnect between having delivered that evidence and requests for more of it puts social scientists in an unenviable position. After all, is the question about social media's potential harm not answered? Wouldn't it be prudent to stop investing resources into researching effects that evidence suggests are not there?

Instead, the field has called for considering necessary nuance in studying social media (Dienlin & Johannes, 2020; Masur, 2021; Meier & Reinecke, 2020; Orben et al., 2020). Many researchers have started to investi-

gate variation around the average association—or lack thereof—linking social media use and well-being. Their aim: to investigate individual social media users and what makes them susceptible to large effects (Aalbers et al., 2021; Beyens et al., 2020; Whitlock & Masur, 2019). Studying variation might be informative, but the field has not yet made explicit the goals behind studying variation—nor has it developed a principled approach to researching or understanding that variation. Before we present such an approach, we believe it is important to pause and ask ourselves: How did we get here? What's the problem? And where do we go from here? Studying variation before a solid understanding of how to answer these questions may investing resources that could be more valuable elsewhere. Here, we tackle those questions in the hope that we can outline a principled approach for the study of social media and well-being.

How did we get here?

Prescribing the impossible

Ever since social media became popular at the turn of the millennium, there have been concerns about their negative effects (e.g. Beard, 2005). While the social sciences were still figuring out how to study social me-

dia, the first popular science books about their harmful effects on our thinking (Carr, 2011) and social fabric (Turkle, 2012) came out and caused a stir in public discourse. Policymakers reacted quickly. Above all, they wanted to know how much time on social media is safe for its users—particularly for children. For example, in 2016, the American Academy of Pediatrics released a policy statement on Media and Young Minds, recommending that children younger than 18 to 24 months avoid digital media (except video-chatting) altogether (Council on Communications and Media, 2016). In other words, the field was tasked to deliver prescriptions, often for individual users: How much screen time is safe for my child?

Such a task might well be impossible. Most work in the quantitative social sciences is traditionally on the group-level, comparing differences between people (Bryan et al., 2021; Hamaker, 2012; Richters, 2021): How are differences in social media use between people related to differences in well-being? Unfortunately, what holds between people on the group-level doesn't automatically translate to members of the group: An effect between people in a large group doesn't mean an effect within each person (see Figure 1a). Such a transfer from between-person to within-person is only admissible under strict assumptions of so-called ergodicity (for details, see Fisher et al., 2018; Hamaker, 2012; Molenaar, 2004). A process is ergodic if it fulfills two conditions (simplified here): First, the psychological structure of a process is the same for each member of the sample (homogeneity). Second, that the structure is stable over time (stationarity; Molenaar and Campbell, 2009). In our case, we'd assume that the relation between social media and well-being is identical for each user and doesn't change over time. Both assumptions are rarely, if ever, fulfilled; behavior, thoughts, and feelings are simply not the same for everyone (Fisher et al., 2018; Gelman, 2015; von Eye, 2009). Put differently: "Individual variability is the rule, not the exception" (Rose et al., 2013, p. 152). Just because across our sample people who use an hour more of social media experience half a point less well-being than those users with an hour less social media doesn't mean an hour more of social media is associated with half a point less well-being for every social media user in the population.

The challenge presented by ergodicity doesn't bode well for social scientists who seek to make prescriptions on social media use. What if we explicitly study within-person effects? We can't generalize from between-person effects to individuals, but if we have repeated measures, can we translate the average within-person effect to each person? We can, but only if relations are

lawful (Haaf & Rouder, 2019; Hamaker, 2012). For example, the relation between media use and well-being might well be positive for everyone, even if there are differences in size. Figure 1b shows an example of an unlawful and Figure 1c shows an example of a lawful process: If relations within each person are not all in the same direction (Figure 1b), the average within-person effect is our best guess for individual media users, but we can't make correct prescriptions for each user. If relations within each person have the same direction (e.g., positive, Figure 1c) prescriptions are possible. Unfortunately, we're not aware of many psychological processes that follow such lawfulness (Bryan et al., 2021; Eronen & Bringmann, 2021; Haaf & Rouder, 2019)—unlike some biochemical processes that will affect almost everyone in a similar way, but to different degrees (Muthukrishna & Henrich, 2019).

Even if we were to believe there's lawfulness to establish in the relation between social media and well-being, that lawfulness rests on several, likely untenable, assumptions. For one, calls for prescriptive recommendations follow a linear dose-response model: How much time on social media will lead to how much change in well-being (Johannes et al., 2022)? From the start, there is good reason to believe such an assumption is overly simplistic (Griffioen et al., 2020; Kaye et al., 2020; Orben et al., 2020; Parry, Fisher, et al., 2021), but it has nonetheless shaped how social scientists studied social media. However, there is no convincing evidence that social media are indeed the dose that causes a response. Most studies in the field cannot make causal claims because there is a lack of theory that would allow to test causal models with cross-sectional data (Eronen & Bringmann, 2021; Rohrer, 2018) or inform us about time-varying confounders in longitudinal data (Hernán, 2018; Rohrer & Murayama, 2021; VanderWeele et al., 2016). The few experiments with high external validity show no causal effect (Hall et al., 2019; Mitev et al., 2021; Przybylski et al., 2021). Therefore, it is questionable whether social media use is indeed the dose that causes a response (i.e., changes in well-being). It also is unclear whether a dose-response relation would be linear, such that with each extra moment of social media use, well-being varies to the same extent as with the next moment (Bruggeman et al., 2019; Johannes et al., 2022; Przybylski & Weinstein, 2017). The most unrealistic assumption might be that of universal effects (Eronen & Bringmann, 2021; von Eye, 2009). Most studies merely look at screen time and its relation to a range of well-being indicators such as life satisfaction or self-esteem. Without taking the motivations of users into account, inspecting different types of content that people engage with on social media, and testing rela-

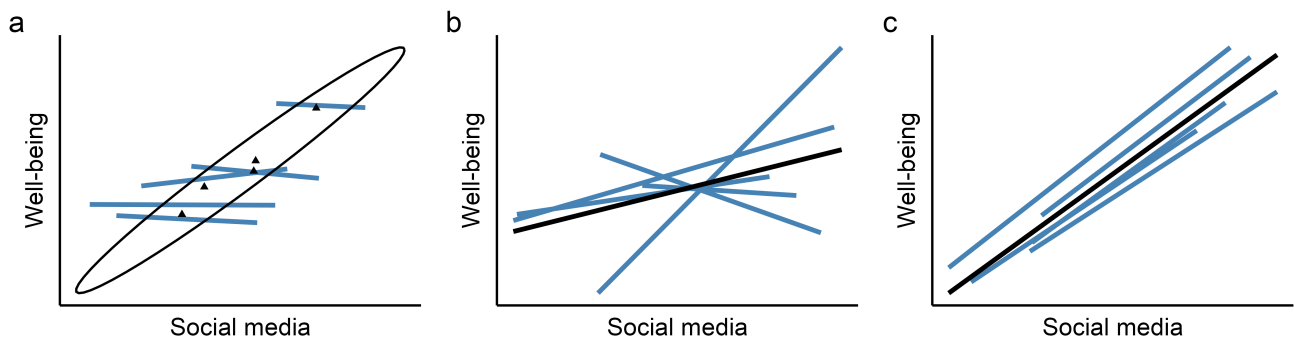


Figure 1

Examples of different between-person and within-person processes in 5 fictitious people. Lines represent regression lines between social media use and well-being. a) No within-person effects (blue lines), but a positive between-person effect (black ellipses, triangles are averages per person on social media use). b) No lawful within-person effect: Even though the average within-person effect (black) is positive, not every participant shows a positive effect (blue). c) Lawful within-person effect: The average within-person effect (black) is positive and can be generalized to every person because every person experiences a positive within-person effect (blue).

tions to mental health, these studies can only give us a coarse understanding of broad, net relations (Bayer et al., 2020; Dienlin & Johannes, 2020; Kross et al., 2021; Kushlev & Leitao, 2020; Orben et al., 2020). The literature is largely ignorant of—or simply hasn't been able to measure (Davidson et al., 2022; Parry, Davidson, et al., 2021; Shaw et al., 2020)—social media use with an adequate level of nuance. Taken together, all those reasons make it close to impossible to make (personalized) prescriptions on social media use: between-person findings generalize to individual users only under strict assumptions that likely don't hold in this field; within-person lawfulness in social media effects is implausible because of a lack of nuance around conceptualizing and measuring social media, motivations, and mental health; the evidence is merely correlational; and we can't be certain we indeed investigate a linear effect.

Dealing with the impossible

For the sake of argument, let's set aside the challenges of making person-specific prescriptions and instead take the current evidence in the literature at face value. Most scholars have reached consensus: Concerns about general social media use don't seem warranted. Large-scale studies don't support the conclusion that there are sizable group-level relations between social media use and well-being (Dickson et al., 2019; Dienlin & Johannes, 2020; Kross et al., 2021; Kushlev & Leitao, 2020; Masur, 2021; Meier et al., *in press*; Meier & Reinecke, 2020; Orben, 2020b; Valkenburg et al., 2022; Whitlock & Masur, 2019)—both on the between-person

and the within-person level (Coyne et al., 2019; Dienlin et al., 2017; Houghton et al., 2018; Orben et al., 2019; Orben & Przybylski, 2019; Schemer et al., 2020; Thorisdottir et al., 2019). In other words, those who use more social media are not worse off compared to those who use them less (between-person) and using social media more than a person usually does is not systematically related to changes in that person's well-being (within-person).

Therefore, we're presented with a curious mismatch: The lack of evidence linking social media and well-being is out of step with public concerns about negative effects, addiction, and distractions (Ellis, 2019; Kardefelt-Winther et al., 2017; Loh & Kanai, 2015; Satchell et al., 2021). When public opinion about media effects is this strong but social scientists are not able to produce evidence in support, maybe they can't be trusted with producing evidence for questions that have a (seemingly) less obvious answer? Such a perception can have serious consequences for research funding and the credibility of social science overall (IJzerman et al., 2020). We believe a widespread acceptance that the average relation between social media use and well-being is negligible may well lead the research area to lose relevance in both academic and public policy discussions.

The field initially reacted to that threat by focusing on subgroups in the population, asking to identify those people who show non-negligible relations between social media use and well-being (Griffioen et al., 2021; Orben, 2020a; Valkenburg & Peter, 2013). This shift from overall, broad relations to subgroups scaffolds between a public who have a wide range of concerns about

the general effects of social media and several specific cases (e.g., radicalization, bullying) that suggest that there are contexts or distinct subgroups for which social media use has large effects (Valkenburg et al., 2021). That is, media effects can differ because of differences in context or differences between people. One proposal focuses on these differences between people and highlights a person-specific approach as a new paradigm for media effects research (Valkenburg, 2022; Valkenburg et al., 2021). The reasoning behind the approach goes as follows: Average relationships between social media use and well-being aren't informative. Instead, all people are different, so we must examine individual social media users and the (within-person) relation between social media use and well-being that is specific to them. Some users will show positive relations, some negative, and others no relation at all, leading to overall negligible average relations (Aalbers et al., 2021; Beyens et al., 2020; Siebers et al., 2021; Valkenburg et al., 2021). Figure 2 (top) illustrates that reasoning: When the average relation between social media use and well-being is zero, but there is variation around the average relation, negative and positive relations 'cancel' each other out. If that variation is large, such a null distribution can hide real harm and benefits. Note that Figure 2 is merely a different way of presenting the information of Figure 1 (b and c panels), showing the distribution of within-person relations that are lawful (all positive, blue distribution, bottom of Figure 2) or not lawful (all other distributions that vary in the sign of the effect). In the past year, several analyses of experience sampling data explored this idea of person-specific media effects and focused on the variation around average relations (Aalbers et al., 2021; Beyens et al., 2020; Siebers et al., 2021; Valkenburg, 2022; Valkenburg et al., 2021).

Such investigations of variation around average relations may be of high value for the field. However, we first need to agree on our goal behind investigating variation (Howard & Hoffman, 2018). We believe that a person-specific approach needs to be clear on that premise to advance the field. Otherwise, prioritizing variation over interpreting and understanding average associations risks atomizing associations. If a person-specific media effects paradigm follows in the footsteps of the person-specific paradigm in Psychology and indeed wants to study media effects that literally are specific to a person, its goals are to make inferences about social media and well-being for individual people, not groups. Alternatively, if it wants to make inferences to a group, its goals are to study variation around average, group-level effects to identify risk factors and susceptible groups, not individuals (Bryan et al., 2021; Howard & Hoffman, 2018). However, it can't conflate

the two approaches and study variation in a group to make person-specific inferences. Put differently: Goal (i.e., person-specific) and methodology (i.e., inferences about individuals) must be aligned. Next, we'll explain how we believe that such a conflation is currently happening. If we decide that we want to study group-level processes, we then need to develop a principled approach towards identifying, interpreting, and explaining variation around average relations.

What's the problem?

There will (likely) always be variation

Because there are so many differences between (e.g., you are different than me) and within people (e.g., you are now different to you earlier), variation around effects is exceptionally likely in nearly any psychological phenomenon (Bryan et al., 2021; Molenaar, 2004; Rose et al., 2013; von Eye, 2009). As far as we know, we have yet to identify an invariant phenomenon in the social sciences. Because human cognition, emotion, and behavior are complex and difficult to measure (Eronen & Bringmann, 2021), it is practically impossible to causally explain them in their totality (Muthukrishna & Henrich, 2019). Consider the well-known Stroop effect: People are slower to name the color of incongruent words (i.e., the word RED in the color blue) compared to congruent words (i.e., the word RED in the color red). We assume the effect follows some lawfulness of cognitive processing that is universal across humans. But it is highly likely that even effects whose direction is universal have variance (Haaf & Rouder, 2019). There are a myriad of different genetic and environmental influences on human behavior—not to speak of the differences in affordances, content, and user motivation for using social media. These influences can and will interact; therefore, each occasion a person uses social media is so multiply determined as to be nearly unique.

As researchers, we're ultimately interested in the individual person (Rose et al., 2013). At the same time, we want to generalize beyond that specific person. These two goals often clash and require different goals and methodologies (Bergman & Vargha, 2013). Quantitative social science currently embraces the nomothetic tradition which aims to make general predictions about the population; it asks what applies in the aggregate. The idiographic tradition aims to make predictions about the individual; it asks what applies in the particular (Hamaker, 2012; Howard & Hoffman, 2018). As we have explained in an earlier section, we can't generalize from the group-level to the individual because of variation between and within people. What applies in the aggregate simply won't apply to the par-

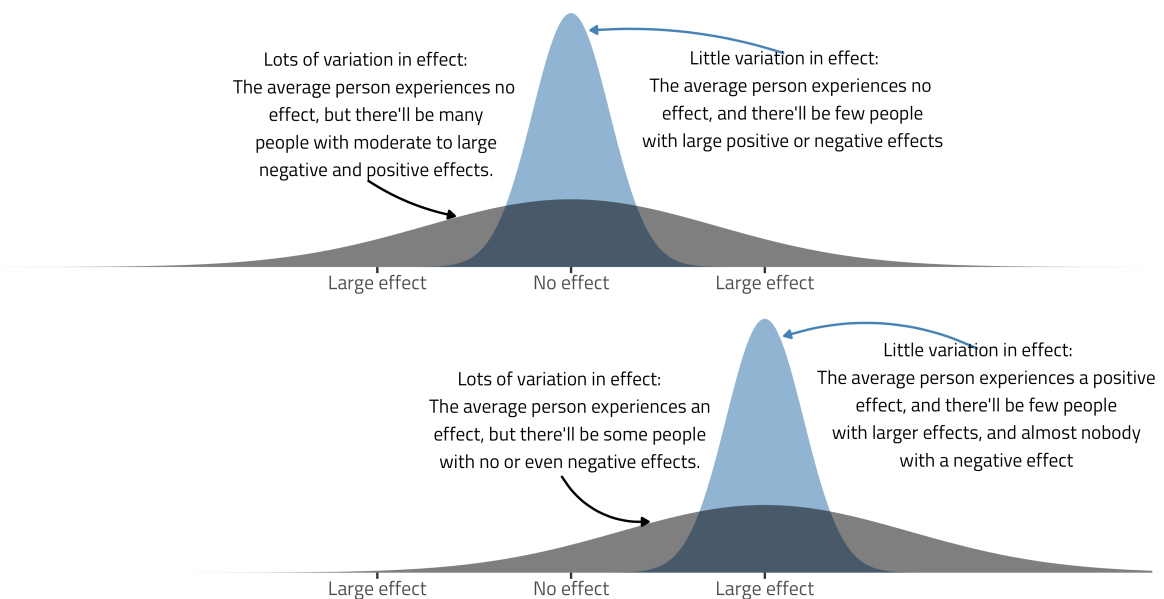


Figure 2

Examples of the distributions of a null average effect (upper panel) and a large average effect (lower panel) under different levels of variation between individuals' effects (little variation in blue and a lot of variation in black).

particular (Molenaar & Campbell, 2009). Studies within the nomothetic tradition often rely on what is called a variable-centered approach: How are variables across people related? For example, when we study how social media relates to well-being in a large group, not taking into account individual people's characteristics, we rely on a variable-centered approach. Studies within the idiographic tradition rely on what is called a person-specific approach: How are variables within this particular person related? For example, when we study how social media relates to well-being for a particular person, including all their individual characteristics, we rely on a person-specific approach.

These traditions require different methodologies (Bergman & Vargha, 2013; Howard & Hoffman, 2018). Let's go back to the Stroop example: A variable-centered approach aggregates over all participants in the Stroop experiment to compare response times of congruent and incongruent words. The result is a group-level estimate that can be generalized to the population the sample is from. A person-specific approach estimates personalized parameters to compare response times within a particular person. The result is a person-specific estimate that can be generalized to the person only (Molenaar, 2013). If we have multiple participants, we run as many models and summarize the results qualitatively—or use newer bottom-up models that find commonalities between people (Beltz et al., 2016). Both

a person-specific and a variable-centered approach will focus on explaining the variation that seems inevitable in human behavior. But they will make inferences on different levels. We see promise in a person-specific approach that investigates individual people and makes inferences to these individual people. However, we believe the field, claiming to follow a person-specific approach, relies on a variable-centered approach instead: estimating and inspecting variation on the group-level. If we accept that not all people are the same and social media effects naturally contain variation, the conclusion that media effects won't be the same for everyone takes the form of a circular argument. Next, we explain how using a variable-centered approach to make person-specific inferences neglects the primary purpose of a variable-centered approach.

Inferential goals and problems

If we adopt a variable-centered approach, we want to study relationships between variables in the population. Say we want to test the relation between social media use and well-being and sample 1,000 people. Next, we build a statistical model that allows us to estimate the direction and magnitude of the relation. If we find that it's negative, we don't want to conclude that the relation only applies to those particular 1,000 people we happened to sample. If we were, we wouldn't need infer-

ential statistics. We could just calculate the size of the relation and have our answer. But we sampled those 1,000 people to draw an inference to the population they come from. Statistical inference within a variable-centered approach is thus necessarily inductive and on the aggregate: To arrive at an inference about the population, we generalize an aggregate estimate from a sample of that population.

Sampling introduces sampling error. Statistical inference attempts to separate signal (i.e., the true effect or association) from noise (i.e., the error), which means there will be variation in our measures—be it caused by measurement error, sampling error, or true variation in the effect. That variation can occur on two levels: Between people (i.e., differences from one person to another) and, if we have multiple measurements per person, within people (i.e., variation around the person's mean). In our statistical model, we should know what sources of variability to account for to identify the signal. Because we want to generalize from the people in our sample to the population, we need to account for variation of people being different from each other. Only if we account for these differences are we allowed to generalize to other people. Social scientists often account for such variation in various forms of mixed-effects models by specifying grouping variables (Bates et al., 2015; Bolker et al., 2009; DeBruine & Barr, 2021)—ideally all sources of variability that we want to generalize over (Yarkoni, 2020). Therefore, when we predict well-being, we obtain fixed (i.e., average, aggregate relation) and random effects (i.e., relations specific to individuals) for social media use. Random slopes mean that the model doesn't assume that the relation will be the same for every participant; the model takes these differences between people into account and provides us with the best estimate of the average relation on the group-level: the fixed effect.

Therefore, fixed effects are the primary outcome of mixed models and we caution against treating them as secondary. For example, next to the fixed effect, Beyens et al. (2020) reported the distribution of random slopes of the relation between social media use and well-being, categorizing individual random slopes according to sign and size. They state: "Because only small subsets of adolescents experience small to moderate changes in well-being, the true effects of social media reported in previous studies have probably been diluted across heterogeneous samples of individuals that differ in their susceptibility to media effects" (Beyens et al., 2020, p. 2). We believe such a conclusion conflates several issues. First, as we've argued earlier, there will (likely) always be variation. Furthermore, for a null relation to result from a treatment that causes roughly equal pro-

portions of negative and positive effects (to cancel each other out) is less likely than merely a true null effect with random variation (Dahly, 2021). Second, focusing on the model's random slopes emphasizes description of the sample over inference to the population. It neglects the purpose of our models: the estimate of the average association in the population. This issue is exacerbated by non-representative samples typically recruited in the field.

That's not to say that the variation around the fixed effect is meaningless, nor that random slopes don't carry information. In fact, a random slope is indeed an estimate specific to that person (Efron & Morris, 1977). However, we first need to agree under what premise we want to study such variation and on which level of inference we operate. If we're after person-specific effects, we need to describe each individual participant and their random slope; but we shouldn't summarize them because such a summary (e.g., '50% of participants had a negative relation') merely describes the sample and thereby defeats the purpose of a person-specific approach. In contrast, under the premise of a variable-centered approach, inspecting variation around the fixed effect complements, but does not in any way supplant, information the fixed effects carry—it can inform us about the expected variation from person to person around the fixed effect. That variation, in turn, can inform us whether we should identify systematic causes for this variation, such as moderators or other predictors of variance. (More on moderators later.) If we agree on that premise and the adequate level of inference, the question is not whether there will be variation around the fixed effect. The questions are rather: How do we estimate variance around the fixed effect? How much variation is there? And how much variance is relevant to warrant further attention? The field must provide benchmarks against which we measure the answers to these questions; it must specify how much variation is meaningful and warrants further investigation. Next, we therefore outline a principled approach to dealing with variation in average relations.

Where do we go from here?

Quantifying variation

How to assess whether there is meaningful variability around the average effect is neither a new challenge nor is it one special to the study of social media. For example, in the field of personalized medicine, there is a heavy debate on how to understand variation in effects and how to demark effects on the individual level from those on the group-level (Senn, 2016, 2018). A similar debate has been going on in the social sciences for sev-

eral years (Fisher et al., 2018; Molenaar, 2004; Molenaar & Campbell, 2009; Richters, 2021). Similarly, Bolger et al., 2019 have addressed the question of meaningful variation in experimental effects extensively and provide an overview of how to deal with effect heterogeneity (i.e., variation in effects) (Hester et al., 2021; Liew et al., 2016). How has media effects research studied variation so far? Researchers most often start with model comparisons, where they compare a model with only a fixed slope (i.e., the effect will be the same for every person) to a model with additional random slopes (i.e., each person will differ to a degree from the overall effect). Another common practice is plotting the distribution of the observed random slopes to demonstrate the variation in the relation between social media use and well-being. A subsequent step is often defining cutoffs for effect sizes following the conventional benchmarks of Cohen, 1988 and describing what proportion of random slopes in the sample exceeds these benchmarks (e.g., 12% of the observed random slopes are considered large). Rather than reinventing the wheel for our area, we aim to integrate work from other fields and translate some steps taken by previous research to a principled approach to study variation in social media research.

To illustrate that approach, we work along an example taken from Beyens et al. (2020) who presented a study on the relation between active and passive social media use and well-being in an experience sampling study. They found a fixed effect for the relation between passive social media use (in steps of five minutes) and well-being (how happy someone felt in the moment on a 7-point Likert scale) of .06. That association was on the within-level: For the average person, spending five more minutes passively on social media in the past hour than they typically do was associated with a 0.06 increase in well-being. That fixed effect was not significant ($p = .440$).

How do we know how much that effect varies between people? The standard deviation of the random slopes provides that answer. In the case of our example, the standard deviation was 0.24 ($\sigma^2 = 0.06$ from Table 3, Model 4b), more than four times as large as the average effect. From the standard deviation, we can calculate an interval around the fixed effect, sometimes referred to as heterogeneity interval (Bolger et al., 2019), as the 2.5 and 97.5 percentiles of the normal distribution implied by the fixed effect and the standard deviation ($0.06 \pm 1.96 \times 0.24$). Therefore, our heterogeneity interval is $[-0.41, 0.53]$. It tells us that 95% of person-specific associations between social media use and well-being in the population would fall within this range. According to the model, some people will experience

negative associations (-0.41) that are 7 times more intense and negative than the average positive association (0.06); others will display positive associations (0.53) that are 9 times larger.

In this example, we used point estimates of the fixed effect and its standard deviation to obtain a heterogeneity interval. In practice, these parameters are estimated from data and therefore introduce their own source of uncertainty that ought to be included in further calculations (e.g., of heterogeneity intervals). Without representations of these uncertainties, for example in the form of posterior distributions, researchers run the risk of making overly confident statements. However, we only had access to point estimates for these examples and therefore continue working with them, while recognizing that in practice such uncertainties should be described.

Note that effect heterogeneity and the uncertainty around the fixed effects are not the same. The fixed effect is the average association between social media use and well-being; its surrounding 95% confidence intervals inform us about variability in that average relation from sample to sample. If we ran infinite studies, 95% of the confidence intervals around the fixed effect would contain the true population average relation. In contrast, the heterogeneity interval informs us about variability in the association from person to person. If we ran infinite studies, 95% of the heterogeneity intervals would contain an individual person's true relation of social media use and well-being. However, the accuracy of these parameters only holds assuming adequate sampling on both the between- and the within-person level. On the between-level, if we sample social media users that are not representative of the population we want to generalize to, our estimate of the variability of the effect is not representative either. The same limitation applies if we don't obtain a representative sample from people's everyday social media use and well-being (e.g., via a random experience sampling procedure). If we don't study a representative sample of a person's life, inferences about the distribution of all participants in our study will be flawed. Therefore, the accuracy of any descriptive analysis of a distribution of individual relations depends on sampling on both levels: the individual and the group.

Assuming adequate sampling, the heterogeneity interval therefore answers exactly the question we are interested in: What social media relations can we expect in the population? Unfortunately, the field has not employed these intervals, which prevents social scientists from being able to quantify variation in media effects from person to person in the population. Merely inspecting random slopes as evidence of meaningful vari-

ation in the relation confounds sample-to-sample variation of the average relation and person-to-person variation around the average relation. We recommend the field adopts the practice of estimating heterogeneity intervals. As a quantitative discipline that is interested in variability of a parameter, we need to define how to estimate that parameter before we can even begin to interpret variability.

Interpreting variation

Now that it can be quantified, should we ignore effect heterogeneity or consider it worthy of further investigation? Many social scientists hold the view that variation around effects is indirect evidence of so-called 'hidden' moderators (Kunert, 2016), thereby seeing all variation as meaningful and worthy of further examination. However, we caution against adopting this position as a default. As we have explained, few, if any, psychological phenomena will be invariant and much variation we can consider noise (e.g., from the sampling strategy, sample size, the size of the fixed effect, measurement error, to name just a few). Explaining all variation may practically be impossible—even within a person-specific modelling approach that sacrifices parsimony of the model for better prediction (Howard & Hoffman, 2018). To distinguish meaningful from random variation—to sort the signal from the noise—we suggest a principled workflow that follows three steps (see Table 1). First, we can compare models as a baseline test. Second, we must define a Region of Practical Equivalence (hereafter ROPE; Kruschke, 2014) around the fixed effect and test our heterogeneity distribution against this ROPE to identify noteworthy variation. Third, we must define a Smallest Effect Size of Interest (hereafter SESOI; Anvari et al., 2021; Lakens et al., 2018) and compare the heterogeneity distribution against it. All of these steps should be taken together, not in a piecemeal way.

First, Bolger et al., 2019 recommend model comparisons as a starting point. During that step, we compare a model without random slopes to a model with random slopes. Goodness of fit is the standard by which model comparisons are judged. If the slopes significantly improve model fit, we have initial evidence that there might be meaningful variation around the average effect. As already outlined earlier, this step is far from conclusive. Theoretically, we know that people are different and a model with random slopes will almost always yield a better fit (Barr et al., 2013). Therefore, model comparison provides a necessary, but not a sufficient, first step.

Second, we must define a ROPE which “indicates a small range of parameter values that are considered to be practically equivalent to [the fixed effect] for the pur-

poses of the particular application” (Kruschke, 2014, p. 336). (Note: We adopted Kruschke’s term; he didn’t apply ROPE to variation.) Let’s apply this definition to our working example. Before we collect data, we decide that our fixed effect of social media use on well-being has noteworthy variation if the effect heterogeneity distribution exceeds a range of ± 0.3 Likert-points around the fixed effect. Note that we operate on the natural scale and not on standardized units because the natural scale is easier to interpret and requires more precise theory (Baguley, 2009). Note also that this number is entirely arbitrary; “ROPE limits, by definition, cannot be uniquely ‘correct,’ but instead are established by practical aims” (Kruschke, 2014, p. 338). We need expert knowledge to determine our ROPE and provide context for analyses which use it as standard in our models.

For some, 0.3 will represent a meaningful and sensible cutoff for this effect; for others, it won’t. Like Bayesian procedures that clearly communicate prior beliefs about an effect, being transparent and putting ROPE up for discussion enables others to better scrutinize how we deal with effect heterogeneity (Dienes, 2019). With this procedure, we communicate to readers that we only find the variation around a fixed effect worthy of further study if that variation doesn’t fall within the ROPE.

After having defined our ROPE, we need to test the variation against the ROPE. Here, we don’t rely on the observed random slopes, but the theoretical distribution around the fixed effect, that is, the heterogeneity distribution from which we draw the heterogeneity interval (Bolger et al., 2019). The observed random effects in the sample distract from the actual purpose of the model, which is to make an inference to the population. As we explained in the section on quantifying variation, we can estimate this theoretical distribution with the fixed effect and its standard deviation.

We then can calculate the area under that theoretical distribution to infer what proportion of media users fall below or above certain thresholds. In our recurring example, we have an average relation of 0.06 and a standard deviation of 0.24. Our ROPE of ± 0.3 Likert-point hence ranges from -0.24 to 0.36 ($.06 \pm 0.3$). Now, we can calculate what proportion of our distribution falls outside the ROPE. For this example, the area outside this range is 22%. Depending on the research context, we could conclude that there is therefore noteworthy variation around the fixed effect (for details on the calculation, see <https://osf.io/b7rpx/>).

Note several points here: Because we use the theoretical distribution, and not observed slopes, we can make an inference to the population. However, as we explained before, we have used a theoretical distribu-

Table 1

An explanation of the three steps of interpreting variation.

Step	Explanation
Model comparison	Statistically compare a model with a fixed effect and random slope to a model with only a fixed effect.
Region of Practical Equivalence (ROPE)	Define a region of practical equivalence and estimate and compare theoretical distribution of average effect against it.
Smalles Effect Size of Interest (SESOI)	Define a smallest effect size of interest and compare theoretical distribution of average effect against it.

tion derived from point estimates of fixed effect and its standard deviation. For an inference that takes uncertainty into account, ideally we need to estimate the proportion of the theoretical distribution outside ROPE for each parameter combination in the posterior. This approach is therefore more informative than merely describing what proportion of observed random slopes are outside a cutoff, because observed random slopes describe the sample, not the population. Second, effect heterogeneity is independent of the location of the fixed effect: We specify the ROPE around wherever the fixed effect will fall. Therefore, ROPE limits (i.e., its width) are relative to the location of the fixed effect.

Third, now that we have tested whether there is considerable variation around the fixed effect, we can move on to investigate the location of the distribution and its width in relation to an absolute limit. This combination answers whether there are meaningful associations in the population. To investigate whether the variation we consider noteworthy also matters practically, we need to define a smallest effect size of interest (SESOI) for the relationship of interest (e.g., social media use and well-being). The SESOI tells us how large an association has to be for us to consider it practically relevant (Anvari et al., 2021; Anvari & Lakens, 2021; Lakens et al., 2018).

Both the ROPE (i.e., width of distribution, relative limits) and the SESOI (i.e., location and width of distribution, absolute limits) matter, see Figure 3: Our effect heterogeneity might well exceed the ROPE, but that doesn't mean it's practically relevant. The distribution of associations we can infer from our sample might well have noteworthy variation, but fall completely within the bounds of our SESOI (see blue distribution in Figure 3). Then we conclude that there is noteworthy variation, but that variation operates within a range we don't consider relevant. On the flipside, our distribution might not exceed the ROPE, but lie completely outside our SESOI (grey distribution in Figure 3). Now we don't find noteworthy variation, but everyone in the population shows a relevant, large enough association. Finally, and probably most common, there may be less clear-

cut cases (red distribution in Figure 3): For example, we might have noteworthy variation, but some parts of the distribution are equivalent to a practically insignificant effect (i.e., inside the SESOI range). Here, we have noteworthy variation and large parts of the population show a large effect.

For our running example, we choose a large SESOI. Just like with ROPE, we define our SESOI on the raw scale. Depending on the outcome, a standardized effect that has medium size by Cohen's (1988) benchmarks might be practically meaningless (Anvari et al., 2021; Baguley, 2009; Lakens et al., 2018). Therefore, once again we need to apply our domain knowledge to define what we consider a meaningful, absolute effect. In our example, say we only regard large associations of at least one Likert-point or larger as relevant (14% on the 7-point response range). We found noteworthy variation in the previous step because the distribution exceeded our ROPE, but only 0.00007% of the distribution fall outside the SESOI (similar to the blue distribution in Figure 3). Theoretically, we can expect 0.00007% of the population to exhibit an association between social media use and well-being that we consider plausibly meaningful. Note again that we're using point estimates; ideally, we inspect what proportion of the heterogeneity interval's lower and upper bounds lie outside the SESOI.

Such a small absolute percentage signals few practically relevant effects in the population. But what if the number had been higher, say 17%? Is 17% enough people to conclude that we need to explain that variation? Again, there is no absolute rule here and the answer depends on the researcher. Some will conclude that the variation in associations in itself is noteworthy and probably worth studying (i.e., ROPE), were it not for the generally small associations (i.e., SESOI): Explaining even noteworthy variation might be inconsequential. Others will conclude that the variation in associations is noteworthy and large enough in enough cases to be relevant and worthy of further study. Whatever researchers decide, we urge them to be explicit and

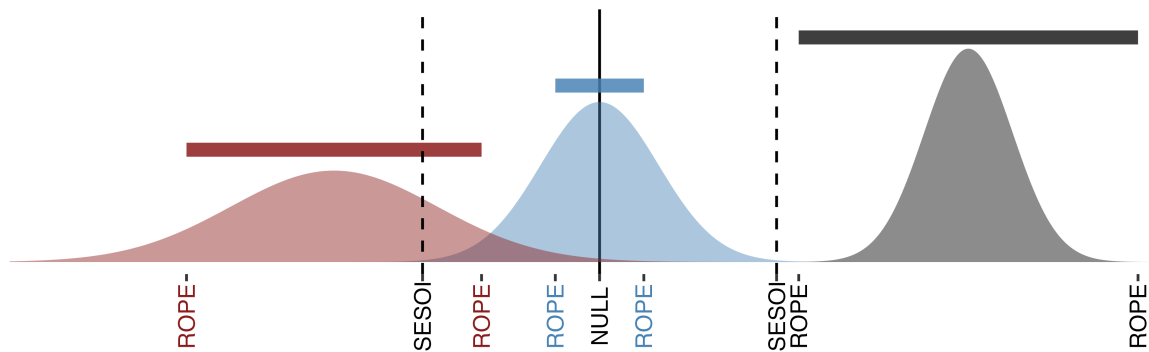


Figure 3

Examples of how ROPE (Region of Practical Equivalence) and SESOI (Smallest Effect Size of Interest) interact. Distributions have different ROPEs (bars on top), but the same SESOI (dashed vertical lines). The red case shows a distribution that is outside ROPE and outside SESOI. The blue case shows a distribution that is outside ROPE, but inside SESOI. The grey case shows a distribution that is inside the ROPE, but outside the SESOI.

transparent in their choices of both ROPE and SESOI. As a minimal standard, we suggest preregistration of ROPE and SESOI as a tool for subjecting our hypothesis of effect heterogeneity to a more severe test (Lakens, 2019)—or display a range of ROPEs and SESOIs so readers can interpret the results better (Dienes, 2019).

Explaining variation

Once we know how to quantify variation in media effects and have determined the magnitude of variation necessary to be relevant for social science research, the final logical step is to ask what factors explain that variation. For whom does the effect differ and for what reasons? A large amount of variation around the average effect can mean that there are unobserved factors that explain why some people show a large and others a small effect. It might be well worth to study these factors. But if we rely on a variable-centered approach for studying those factors, we are yet again interested in fixed effects. We agree with previous research that it's an important step for social media research to identify those people who are more susceptible to media effects (Beyens et al., 2020; Griffioen et al., 2021; Orben, 2020b; Orben et al., 2019; Valkenburg & Peter, 2013). However, as long as we're committed to group-level inferences, that step should not be taken for individual social media users; instead, we must study systematic individual differences or differences in the content of social media that can account for variation in social media effects.

Identifying susceptible people means identifying factors that can explain systematic variation in the effect. Statistically, those factors are modelled as moderators (Bryan et al., 2021). Moderators explain variation in

the effect across the population, because they model how our average effect differs, on average, between groups of people. Once more, consider the Stroop effect that we used as an example earlier on. The fixed effect will show that people, on average, are slower on incongruent trials compared to congruent trials. However, that effect likely varies, such that some people show little slowing and others extreme slowing. For example, differences in visual acuity might induce systematic differences between participants. If some participants have forgotten their contact lenses, they might be slower to read and therefore show a different effect.

In the case of social media and well-being, if we find that the relation between social media use and well-being has high variability, it's possible that modelling knowledge about group membership can explain parts of that variability. For example, whether someone identifies with a particular gender might be a moderator because the relation is present for teenage girls, but absent for boys (Orben et al., 2019). But note that we infer that this moderation generalizes only to the population which was sampled: A large group of British young people aged 10 to 15 years old. We're not saying the relation is negative for a specific girl, or null for a specific boy. There's little doubt many girls in the data show no relation whereas a number of boys show negative associations. Put differently, identifying moderators will reduce effect heterogeneity, but cannot entirely eliminate it (Bolger et al., 2019). Identifying moderators echoes calls to take effect heterogeneity seriously (Bryan et al., 2021), because moderators answer the question of what factors in groups of people can explain media effects for groups—not what factors play a role for individual media users.

Identifying moderators in a disciplined, theory-driven, and accurate way is difficult and we expect social scientists will be tempted to adopt a 'shot-gun approach' and measure and test a large number of constructs as moderators. This strategy is doomed to failure. It will lead to high false positive rates and fool social scientists into giving more weight to those moderators that 'worked' (Munafò et al., 2017; Nosek et al., 2018). Testing a wide slate of seemingly plausible moderators will inevitably yield statistically significant results; but ignoring researcher degrees of freedom means these results will not be informative unless more advanced statistical methods are used (Gelman & Loken, 2013; Simmons et al., 2011). Such exploratory findings must be subjected to confirmatory replications (Frankenhuis & Nettle, 2018). Ideally, theory should identify moderators that researchers test in truly confirmatory tests (Fried, 2020). Only such an approach can systematically explain effect heterogeneity that can be generalized to the population.

If, instead, social media research is truly interested in following a person-specific approach and studying effects for each individual person, we argue that such a focus requires a different approach. Researchers then need to study a single person over many measurement occasions, which is sometimes called $N = 1$ studies. They represent an intriguing alternative research direction and are gaining popularity (Matias et al., 2022). Such $N = 1$ studies allow inferences to the person under study only; they reveal effects unique to the specific person. Given the noisiness of social behavior, obtaining a representative sample of that specific person's usage episodes as well as ensuring enough power to detect a potentially small effect specific to that person may require measuring variables hundreds of times and different analysis techniques (e.g., p-techniques; Molenaar and Campbell, 2009; Rose et al., 2013).

Once we've conducted a model per person, we can informally summarize the results across all people's models, but it's inadmissible to aggregate the results to generalize to the population these people come from. Recent statistical models address this problem by employing a bottom-up approach and finding commonalities among all person-specific models, whilst still allowing each person their own model with their own parameters (GIMME, Beltz et al., 2016).

Alternatively, researchers can also describe and understand an individual person through a qualitative approach. A qualitative approach won't lead to quantifiable social media effect estimates, but to a nuanced understanding of social media use and well-being in that specific person.

Regardless of whether it's our goal to make inferences

to the group or to an individual, identifying and testing moderators should only happen in connection with testing causality. A common cause (aka confounder) might increase variation in the relation between two variables that are truly unrelated. However, it would be a fallacy to affirm the consequent: Finding that controlling for a third variable reduces heterogeneity doesn't mean this variable is a common cause. Similarly, finding that a third variable moderates a relation doesn't necessarily mean that we've identified a causal process (Rohrer et al., 2021). Introducing a time lag and focusing on within-person moderators can help in identifying such causality, but it's no guarantee that we've truly identified causality (Rohrer & Murayama, 2021; VanderWeele et al., 2016). For that, we need careful experimentation, causal modelling, and stronger theories (Rohrer, 2018). The relation between aggregate social media use and well-being we currently have can only be a stand-in for a causal effect, and the field needs to put more effort into understanding causes and effects. Studying variation in average relations is thus only a means to the end of identifying causality.

Conclusion

Social science has shown that the average relations linking general social media use to general well-being are mostly close to zero. We can use these average relations as a vantage point to identify factors that make some people more or less susceptible to potential effects of social media. However, we believe the field must be clear on what inferences we want to make if we pivot from studying groups to studying individuals. Do we want to follow a person-specific approach? If so, we must inspect individual people and make inferences to individual people. Or do we want to inspect variation on the group-level? Then we must inspect effect heterogeneity on the group-level, and not conflate this process with an idiographic tradition.

Here, we've argued that the field aims at individual-level inferences, but inspects group-level variation. We've outlined how social media effects research has gotten to this state and propose a path forward. Either we commit to group-level inferences which follow a principled approach to the study of effect heterogeneity: continue investigating fixed effects, develop principles to quantify and interpret effect heterogeneity, and identify moderators of the relation between social media use and well-being. Or we commit to a person-specific approach and focus on making inferences to individual social media users and conduct more qualitative and $N = 1$ studies. Either of these paths will be insightful, but we mustn't confuse them.

Author Contact

Niklas Johannes <https://orcid.org/0000-0001-6612-2842>
 Philipp K. Masur <https://orcid.org/0000-0003-3065-7305>
 Matti Vuorre <https://orcid.org/0000-0001-5052-066X>
 Corresponding author: Andrew K Przybylski
<https://orcid.org/0000-0001-5547-2185>,
 andy.przybylski@oii.ox.ac.uk

Conflict of Interest and Funding

Grants from the Huo Family Foundation and Economic and Social Research Council (ES/T008709/1) supported NJ, MV, and AKP. The authors declare no conflicts of interest. The funders had no role in the conception of this paper, decision to publish, or preparation of the manuscript.

Author Contributions

Niklas had the idea for this article and wrote a first draft. He discussed that draft with Philipp, Andy, and Matti. Andy gave feedback on the structure of the article in its early stages; Matti provided feedback on the logic of effect heterogeneity; Philipp gave feedback on structure and co-wrote the next drafts with Niklas, including developing the principled approach of interpreting variation. Niklas created the figures in dialogue with Philipp and Andy first; then Matti and Philipp revised several of the figures. All authors reviewed, revised, and approved the final version of the manuscript.

Open Science Practices



The editorial process for this article relied on streamlined peer review where peer reviews obtained from previous journal(s) were moved forward and used as the basis for the editorial decision. These reviews are shared in the supplementary files. The identities of the reviewers are shown or hidden in accordance with the policy of the journal that originally obtained them.

This article earned the Open Code badge for making the analysis code available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process is published in the online supplement.

References

- Aalbers, G., Vanden Abeele, M. M. P., Hendrickson, A. T., de Marez, L., & Keijsers, L. (2021). Caught in the moment: Are there person-specific associations between momentary procrastination and passively measured smartphone use? *Mobile Media & Communication*, 2050157921993896. <https://doi.org/10.1177/2050157921993896>
- Anvari, F., Kievit, R., Lakens, D., Przybylski, A. K., Tiokhin, L., Wiernik, B. M., & Orben, A. (2021). Evaluating the practical relevance of observed effect sizes in psychological research. *PsyArXiv*. <https://doi.org/10.31234/osf.io/g3vtr>
- Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology*, 96, 104159. <https://doi.org/10.1016/j.jesp.2021.104159>
- Appel, M., Marker, C., & Gnambs, T. (2020). Are social media ruining our lives? a review of meta-analytic evidence. *Review of General Psychology*, 24(1), 60–74. <https://doi.org/10.1177/1089268019880891>
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100(3), 603–617. <https://doi.org/10.1348/000712608X377117>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bayer, J. B., Triu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual Review of Psychology*, 71(1), 471–497. <https://doi.org/10.1146/annurev-psych-010419-050944>
- BBC. (2021). China cuts children's online gaming to one hour. *BBC News*. <https://www.bbc.com/news/technology-58384457>
- Beard, K. W. (2005). Internet addiction: A review of current assessment techniques and potential assessment questions. *CyberPsychology & Behavior*, 8(1), 7–14. <https://doi.org/10.1089/cpb.2005.8.7>
- Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Moleenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analy-

- sis of clinical data. *Assessment*, 23(4), 447–458. <https://doi.org/10.1177/1073191116648209>
- Bergman, L. R., & Vargha, A. (2013). Matching method to problem: A developmental science perspective. *European Journal of Developmental Psychology*, 10(1), 9–28. <https://doi.org/10.1080/17405629.2012.732920>
- Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, 10(1), 10763. <https://doi.org/10.1038/s41598-020-67727-7>
- Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, 148(4), 601–618. <https://doi.org/10.1037/xge0000558>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Bruggeman, H., Van Hiel, A., Van Hal, G., & Van Dongen, S. (2019). Does the use of digital media affect psychological well-being? an empirical test among children aged 9 to 12. *Computers in Human Behavior*, 101, 104–113. <https://doi.org/10.1016/j.chb.2019.07.015>
- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980–989. <https://doi.org/10.1038/s41562-021-01143-3>
- Carr, N. (2011). *The shallows—how the internet is changing the way we think, read and remember*. Atlantic Books.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Lawrence Erlbaum.
- Council on Communications and Media. (2016). Media and young minds. *Pediatrics*, 138(5). <https://doi.org/10.1542/peds.2016-2591>
- Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2019). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, 106160. <https://doi.org/10.1016/j.chb.2019.106160>
- Dahly, D. (2021). Perils of personalized treatment effects. <https://statsepi.substack.com/p/perils-of-personalized-treatments>
- Davidson, B. I., Ellis, D. A., Stachl, C., Taylor, P. J., & Joinson, A. N. (2022). Measurement practices exacerbate the generalizability crisis: Novel digital measures can help. *Behavioral and Brain Sciences*, 45. <https://doi.org/10.1017/S0140525X21000534>
- DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), 1–15. <https://doi.org/10.1177/2515245920965119>
- Dickson, K., Richardson, M., Kwan, I., Macdowall, W., Burchett, H., Stansfield, C., & Thomas, J. (2019). *Screen-based activities and children and young people's mental health and psychosocial wellbeing: A systematic map of reviews*. EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Dienes, Z. (2019). How do i know what my theory predicts? *Advances in Methods and Practices in Psychological Science*, 2(4), 364–377. <https://doi.org/10.1177/2515245919876960>
- Dienlin, T., & Johannes, N. (2020). The impact of digital technology use on adolescent well-being. *Dialogues in Clinical Neuroscience*, 22(2), 135–142. <https://doi.org/10.31887/DCNS.2020.22.2/tdienlin>
- Dienlin, T., Masur, P. K., & Trepte, S. (2017). Reinforcement or displacement? the reciprocity of ftf, im, and sns communication and their effects on loneliness and life satisfaction. *Journal of Computer-Mediated Communication*, 22(2), 71–87. <https://doi.org/10.1111/jcc4.12183>
- Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236(5), 119–127.
- Ellis, D. A. (2019). Are smartphones really that bad? improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.03.006>
- Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, 16(4), 779–788. <https://doi.org/10.1177/1745691620970586>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>
- Frankenhuis, W., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives*

- on *Psychological Science*, 13(4), 439–447. <https://doi.org/10.1177/1745691618767878>
- Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31(4), 271–288. <https://doi.org/10.1080/1047840X.2020.1853461>
- Gelman, A. (2015). The connection between varying treatment effects and the crisis of unreplicable research: A bayesian perspective. *Journal of Management*, 41(2), 632–643. <https://doi.org/10.1073/pnas.1711978115>
- Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, 348(1-17), 3.
- Griffioen, N., Rooij, M. v., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, 1(1). <https://doi.org/10.1037/tmb0000005>
- Griffioen, N., Scholten, H., Lichtwarck-Aschoff, A., van Rooij, M., & Granic, I. (2021). Everyone does it—differently: A window into emerging adults’ smartphone use. *Humanities and Social Sciences Communications*, 8(1), 1–11. <https://doi.org/10.1057/s41599-021-00863-1>
- Grimes, T., Anderson, J. A., & Bergen, L. (2008). Media violence and aggression: Science and ideology.
- Haaf, J. M., & Rouder, J. N. (2019). Some do and some don’t? accounting for variability of individual difference structures. *Psychonomic Bulletin & Review*, 26(3), 772–789. <https://doi.org/10.3758/s13423-018-1522-x>
- Hall, J. A., Xing, C., Ross, E. M., & Johnson, R. M. (2019). Experimentally manipulating social media abstinence: Results of a four-week diary study. *Media Psychology*, 0(0), 1–17. <https://doi.org/10.1080/15213269.2019.1688171>
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). Guilford Press.
- Hernán, M. A. (2018). The c-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, 108(5), 616–619. <https://doi.org/10.2105/AJPH.2018.304337>
- Hester, N., Xie, S. Y., & Hehman, E. (2021). Little between-region and between-country variance when people form impressions of others. *Psychological Science*, 1–11. <https://doi.org/10.1177/09567976211019950>
- Houghton, S., Lawrence, D., Hunter, S. C., Rosenberg, M., Zadow, C., Wood, L., & Shilton, T. (2018). Reciprocal relationships between trajectories of depressive symptoms and screen media use during adolescence. *Journal of Youth and Adolescence*, 47(11), 2453–2467. <https://doi.org/10.1007/s10964-018-0901-y>
- Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, 21(4), 846–876. <https://doi.org/10.1177/1094428117744021>
- IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 1–3. <https://doi.org/10.1038/s41562-020-00990-w>
- Johannes, N., Dienlin, T., Bakhshi, H., & Przybylski, A. K. (2022). No effect of different types of media on well-being. *Scientific Reports*, 12(61), 1–13. <https://doi.org/10.1038/s41598-021-03218-7>
- Kardefelt-Winther, D., Heeren, A., Schimmenti, A., van Rooij, A., Maurage, P., Carras, M., Edman, J., Blaszczynski, A., Khazaal, Y., & Billieux, J. (2017). How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction*, 112(10), 1709–1715. <https://doi.org/10.1111/add.13763>
- Kaye, L. K., Orben, A., Ellis, D. A., Hunter, S. C., & Houghton, S. (2020). The conceptual and methodological mayhem of “screen-time”. *Open Science Framework*. <https://doi.org/10.31219/osf.io/u4hqh>
- Kross, E., Verduyn, P., Sheppes, G., Costello, C. K., Jonides, J., & Ybarra, O. (2021). Social media and well-being: Pitfalls, progress, and next steps. *Trends in Cognitive Sciences*, 25(1), 55–66. <https://doi.org/10.1016/j.tics.2020.10.005>
- Kruschke, J. K. (2014). *Doing bayesian data analysis: A tutorial introduction with r (2nd edition)*. Academic Press.
- Kunert, R. (2016). Internal conceptual replications do not increase independent replication success. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-016-1030-9>

- Kushlev, K., & Leitaó, M. R. (2020). The effects of smartphones on well-being: Theoretical integration and research agenda. *Current Opinion in Psychology*, 36, 77–82. <https://doi.org/10.1016/j.copsyc.2020.05.001>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259–269. <https://doi.org/10.1177/2515245918770963>
- Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, 23(5), 1639–1646. <https://doi.org/10.3758/s13423-016-1032-7>
- Loh, K. K., & Kanai, R. (2015). How has the internet reshaped human cognition? *The Neuroscientist*. <https://doi.org/10.1177/1073858415595005>
- Masur, P. K. (2021). Digital communication effects on loneliness and life satisfaction. In *Oxford research encyclopedia of communication*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228613.013.1129>
- Matias, J. N., Pennington, E., & Chan, Z. T. (2022). Testing concerns about technology's behavioral impacts with N-of-one trials. <https://osf.io/jtdng/>
- Meier, A., Domahidi, E., & Günter, E. (in press). Computer-mediated communication and mental health: A computational scoping review of an interdisciplinary field. In S. Yates & R. E. Rice (Eds.), *The oxford handbook of digital technology and society*. Oxford University Press.
- Meier, A., & Reinecke, L. (2020). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. *Communication Research*, 1–28. <https://doi.org/10.1177/0093650220958224>
- Mitev, K., Weinstein, N., Karabeliova, S., Nguyen, T., Law, W., & Przybylski, A. (2021). Social media use only helps, and does not harm, daily interactions and well-being. *Technology, Mind, and Behavior*. <https://doi.org/10.1037/tmb0000033>
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar, P. C. M. (2013). On the necessity to use person-specific data analysis approaches in psychology. *European Journal of Developmental Psychology*, 10(1), 29–39. <https://doi.org/10.1080/17405629.2012.747435>
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., Simonsohn, U., & Wagenmakers, E.-J. (2017). A manifesto for reproducible science. *Human Nature Behavior*, 1, 1–9. <https://doi.org/10.1038/s41562-016-0021>
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-018-0522-1>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry*, 61(3), 336–348. <https://doi.org/10.1111/jcpp.13190>
- Orben, A. (2020a). The sisyphian cycle of technology panics. *Perspectives on Psychological Science*, 15(5), 1143–1157. <https://doi.org/10.1177/1745691620919372>
- Orben, A. (2020b). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*. <https://doi.org/10.1007/s00127-019-01825-4>
- Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences*, 116(21), 10226–10228. <https://doi.org/10.1073/pnas.1902058116>
- Orben, A., & Przybylski, A. K. (2019). Screens, teens, and psychological well-being: Evidence from three time-use-diary studies. *Psychological Science*, 30(5), 682–696. <https://doi.org/10.1177/0956797619830329>
- Orben, A., Weinstein, N., & Przybylski, A. K. (2020). Only holistic and iterative change will fix digital technology research. *Psychological Inquiry*, 31(3), 235–241. <https://doi.org/10.1080/1047840X.2020.1820221>
- Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digi-

- tal media use. *Nature Human Behaviour*, 1–13. <https://doi.org/10.1038/s41562-021-01117-5>
- Parry, D. A., Fisher, J. T., Mieczkowski, H., Sewall, C. J. R., & Davidson, B. I. (2021). Social media and well-being: A methodological perspective. *Current Opinion in Psychology*. <https://doi.org/10.1016/j.copsyc.2021.11.005>
- Przybylski, A. K., Nguyen, T. T., Law, W., & Weinstein, N. (2021). Does taking a short break from social media have a positive effect on well-being? evidence from three preregistered field experiments. *Journal of Technology in Behavioral Science*. <https://doi.org/10.1007/s41347-020-00189-w>
- Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the goldilocks hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science*, 28(2), 204–215. <https://doi.org/10.1177/0956797616678438>
- Richters, J. E. (2021). Incredible utility: The lost causes and causal debris of psychological science. *Basic and Applied Social Psychology*, 43(6), 366–405. <https://doi.org/10.1080/01973533.2021.1979003>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2021). That’s a lot to PROCESS! pitfalls of popular path models. *PsyArXiv*. <https://doi.org/10.31234/osf.io/paeb7>
- Rohrer, J. M., & Murayama, K. (2021). These are not the effects you are looking for: Causality and the within-/between-person distinction in longitudinal data analysis. *PsyArXiv*. <https://doi.org/10.31234/osf.io/tg4vj>
- Rose, L. T., Rouhani, P., & Fischer, K. W. (2013). The science of the individual. *Mind, Brain, and Education*, 7(3), 152–158. <https://doi.org/10.1111/mbe.12021>
- Satchell, L. P., Fido, D., Harper, C. A., Shaw, H., Davidson, B., Ellis, D. A., Hart, C. M., Jalil, R., Bartoli, A. J., Kaye, L. K., Lancaster, G. L. J., & Pavetich, M. (2021). Development of an offline-friend addiction questionnaire (O-FAQ): Are most people really social addicts? *Behavior Research Methods*, 53(3), 1097–1106. <https://doi.org/10.3758/s13428-020-01462-9>
- Schemer, C., Masur, P. K., Geiß, S., Müller, P., & Schäfer, S. (2020). The impact of internet and social media use on well-being: A longitudinal analysis of adolescents across nine years. *Journal of Computer-Mediated Communication*. <https://doi.org/10.1093/jcmc/zmaa014>
- Senn, S. (2016). Mastering variation: Variance components and personalised medicine. *Statistics in Medicine*, 35(7), 966–977. <https://doi.org/10.1002/sim.6739>
- Senn, S. (2018). Statistical pitfalls of personalized medicine. *Nature*, 563(7733), 619–621. <https://doi.org/10.1038/d41586-018-07535-2>
- Shaw, H., Ellis, D. A., Geyer, K., Davidson, B. I., Ziegler, F. V., & Smith, A. (2020). Quantifying smartphone “use”: Choice of measurement impacts relationships between “usage” and health. *Technology, Mind, and Behavior*, 1(2). <https://doi.org/10.1037/tmb0000022>
- Siebers, T., Beyens, I., Pouwels, J. L., & Valkenburg, P. M. (2021). Social media and distraction: An experience sampling study among adolescents. *Media Psychology*, 1–24. <https://doi.org/10.1080/15213269.2021.1959350>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Thorisdottir, I. E., Sigurvinsdottir, R., Asgeirsdottir, B. B., Allegrante, J. P., & Sigfusdottir, I. D. (2019). Active and passive social media use and symptoms of anxiety and depressed mood among icelandic adolescents. *Cyberpsychology, Behavior, and Social Networking*. <https://doi.org/10.1089/cyber.2019.0079>
- Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic books.
- Valkenburg, P. M. (2022). Social media use and well-being: What we know and what we need to know. *Current Opinion in Psychology*, 45, 101294. <https://doi.org/10.1016/j.copsyc.2021.12.006>
- Valkenburg, P. M., Meier, A., & Beyens, I. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology*, 44, 58–68. <https://doi.org/10.1016/j.copsyc.2021.08.017>
- Valkenburg, P. M., & Peter, J. (2013). The differential susceptibility to media effects model. *Journal of Communication*, 63(2), 221–243. <https://doi.org/10.1111/jcom.12024>

- Valkenburg, P. M., Pouwels, J. L., Beyens, I., van Driel, I. I., & Keijsers, L. (2021). Adolescents' social media experiences and their self-esteem: A person-specific susceptibility perspective. *Technology, Mind, and Behavior*. <https://doi.org/10.1037/tmb0000037>
- vanden Abeele, M. M. P. (2020). Digital wellbeing as a dynamic construct. *Communication Theory*. <https://doi.org/10.1093/ct/qtaa024>
- VanderWeele, T. J., Jackson, J. W., & Li, S. (2016). Causal inference and longitudinal data: A case study of religion and mental health. *Social Psychiatry and Psychiatric Epidemiology*, 51(11), 1457–1466. <https://doi.org/10.1007/s00127-016-1281-9>
- von Eye, A. (2009). Universals and individuals—is this the end of the discussion? *Measurement: Interdisciplinary Research & Perspective*, 7(1), 3–7. <https://doi.org/10.1080/15366360802715320>
- Whitlock, J., & Masur, P. K. (2019). Disentangling the association of screen time with developmental outcomes and well-being: Problems, challenges, and opportunities. *JAMA Pediatrics*. <https://doi.org/10.1001/jamapediatrics.2019.3191>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. <https://doi.org/10.1017/S0140525X20001685>