nature reviews psychology

https://doi.org/10.1038/s44159-026-00603-9

Perspective

Check for updates

Opportunities and risks of big data for the methods and culture of psychological science

Gabrielle C. Ibasco [1] ✉, Friedrich M. Götz [1], Luke Clark [1,2] & Jessica L. Tracy [1]

Abstract

The use of big data, such as large-scale observational and/or archival data, can help psychology to overcome problems with generalizability and enable the study of real-world behaviours. At the same time, big data can be overused or misapplied, which risks reintroducing issues such as an over-reliance on correlations bereft of explanatory power. It is therefore essential that psychologists consider how to leverage big data in an informed and intentional manner. In this Perspective, we describe not only the methodological contributions and limitations of big data, but also the cultural opportunities and risks these approaches evoke; that is, how big data shapes the ways in which psychologists conceptualize and tackle research questions. Of particular concern is that an increased emphasis on big data in combination with academia's publish-or-perish incentive structure could shift the field from 'theories-guide-data' to 'data-guide-theories'. To address this concern and others, we discuss how big data might be harmonized with experimental approaches, which would enable scholars to harness the power of big data while avoiding its pitfalls.

Sections

[1]Department of Psychology, The University of British Columbia, Vancouver, British Columbia, Canada. [2]Djavad Mowafaghian Centre for Brain Health, The University of British Columbia, Vancouver, British Columbia, Canada. ✉e-mail: gcibasco@gmail.com

Introduction

Advances in how large amounts of information are digitally stored and accessed has ushered in an era of big data in psychology, a field that was largely dominated by small-scale lab studies throughout the twentieth century. Specifically, traditional psychological studies tend to feature sample sizes of dozens or hundreds of participants, a limited set of variables, and longitudinal observations spread across a few coarse timeframes. By contrast, big-data projects might involve tens or hundreds of thousands or even millions of participants (big *N*)[1], leverage hundreds of predictors and dependent variables (big *V*)[2,3], and consider highly granular temporal data across both short intervals (daily, hourly or minute-by-minute assessments)[4] and longer time frames spanning decades or centuries (big *T*)[5,6] (for overviews of the types of big-data projects in psychology, see refs. 7,8).

The emergence of big data in psychology has been a gradual process over the past two decades. Several of its first applications can be traced to online surveys that offered instant, anonymous and free results to participants. For example, the Gosling–Potter Internet Personality Project, a dataset collected through a public online survey called The Big Five Project Personality Test (*N* = 361,703 in 2004, ref. 9; *N* > 10,000,000 as of 2026, ref. 10) constitutes the largest known dataset on the Big Five personality traits and spawned research examining differences in personality traits[11] and self-esteem[12] across the lifespan. In the 2010s, the rise of social media and online commerce prompted the use of digital footprints[13] such as social media engagement[14,15] and ad clicks[16] as behavioural indicators of psychological constructs.

As large-scale datasets have become more accessible in psychology, so have a suite of advanced data-science approaches to analyse them. For example, machine learning algorithms in which predictive models are trained on large-scale archival data (such as medical records) have been used to predict group classifications (such as mental health diagnoses[17,18]). Moreover, innovations in natural language processing[19] have enabled researchers to automatically annotate large corpora of text data for purposes beyond classification, such as tracking cultural change over time[20,21], and have more generally enabled researchers to study text and speech as linguistic markers of the human mind[22–25]. Common to all these manifestations of big data is their observational nature: researchers publishing papers using these datasets were usually not involved in manipulating variables or developing assessment tools or surveys. Big-data studies are likely to continue being influential moving forward because, unlike traditional laboratory approaches, they enable psychologists to more efficiently examine generalizable trends and patterns.

In this Perspective, we offer a balanced view on the contributions and limitations of big data to psychological science. Prior reviews tended to either laud big data as the dawn of a new age in research[26] or focus heavily on limitations in common applications of big-data methods for drawing inferences[27]. Here, we zoom out from viewing big data as a methodological approach to also consider its cultural implications for psychological science; that is, the kinds of questions researchers ask and the kinds of answers they deem satisfactory. First, we discuss the methodological advantages of big-data approaches and the broader opportunities of big data for how psychological science is conducted. Next, we consider the limitations of big data methodologically and risks for the research culture of psychological science. We conclude by proposing a way forward for psychology and describe hallmark examples of how big data could be harmonized with traditional laboratory-based experiments to promote a more holistic, robust science.

The scope of 'big data' can be extremely broad, and we do not believe that the label can or should be understood in reference to one singular category. Here, our arguments will largely focus on one of the most prototypical and intuitive use-cases of big data[7,28]: large-scale observational (archival or secondary) datasets. These are datasets that, at a minimum, involve big *N* (although they commonly also feature big *V* and/or big *T;* Box 1) and, importantly, do not involve direct manipulation or intervention by the researchers conducting analyses. Prior reviews and guidelines have resisted identifying numerical cut offs for what counts as big data[7,29]. Indeed, cut offs have been critiqued in other domains in psychology. For example, thresholds for what is considered a large versus small Cohen's *d* are highly subjective and vary across research topics in psychology[30,31]. We believe the same nuance should be applied when conceptualizing big data. Thus, 'large-scale' is left to interpretation and often depends on context, but it generally entails a sample size beyond what a single laboratory can feasibly collect for a given research topic and question. Finally, the terms 'big data,' 'data science' and 'computational modelling' are often used interchangeably in discussions of big-data approaches[26,32,33]. In this Perspective, we distinguish large-sale datasets themselves (big data) from the quantitative approaches used to analyse them (data science and computational modelling; Box 1).

Box 1 | Key terms

Big data
Large-scale datasets that exceed traditional lab studies in their size and/or scope, such that they are beyond what a single laboratory can typically collect in one data collection effort. In psychology, these datasets often take the form of secondary or archival observational datasets for public or proprietary use.

Big *N*
A type of big-data study featuring a large number of observations beyond what a single laboratory can typically collect in one data collection effort, often in the range of tens of thousands or more.

Big *V*
A type of big-data study featuring a large number of measured variables.

Big *T*
A type of big-data study featuring a large number of timepoints for measurement.

Data science
A broad set of principles and tools guiding the extraction of relevant knowledge from big data.

Computational modelling
A data-science approach that involves fitting algorithmic models of cognitive or social processes to data to evaluate competing theories about how behaviour is produced. Parameters for these models can be purely theory driven (based on a priori assumptions) and/or data driven (fine tuned on the basis of real-world patterns observed in big data).

Opportunities of big data

Many psychologists have adopted big data approaches because they enable statistical inferences that can help to overcome the limitations of traditional, laboratory-based research. More broadly, the use of big data might improve how psychologists ask and evaluate research questions with respect to phenomena in the real world.

Methodological advantages

Underpowered sample sizes have been a pervasive problem in traditional psychology studies because they increase the likelihood of false-positive errors and inflate effect size estimates[34,35]. Although small sample sizes are not necessarily inherent to laboratory studies, collecting adequately sized samples in the laboratory is more challenging and costly than accessing large-scale datasets on a public online repository. With big data, researchers can access thousands or even millions of observations at little to no financial cost (see, for example, the World Values Survey and European Social Survey), which circumvents concerns about statistical power and enables precise inferences about small but meaningful effects[36].

In biological psychology and adjacent fields such as developmental neuroscience, laboratory studies are necessary for acquiring physiological measurements. However, scholars have noted the consequences of low statistical power in these fields, such as inflated estimates of correlations in early neuroimaging work[35,37]. To address these limitations, big *N* data collection has been made possible through government-funded initiatives such as the UK Biobank (a comprehensive, prospective biomedical dataset[38]) as well as research consortia that leverage testing across multiple research sites. For example, the Adolescent Brain Cognitive Development (ABCD) study harnesses data originally collected across 21 different sites to assess large-scale, longitudinal associations between brain activity and behaviour[39]. Importantly, these datasets have been made available for use by researchers other than those who collected them, enabling secondary analyses.

Another major benefit of big-data approaches is that they enable researchers to test predictions about how constructs operate at different levels of analysis. This benefit is particularly germane to social-personality psychology, in which theories often stress the embeddedness of individuals within contexts, but which has historically relied on methods constrained to the individual level of analysis[40]. Large-scale studies using multilevel analyses can circumvent this limitation. For example, one study found that regional differences in personality predicted individual spending habits (assessed using real-world behavioural spending data from 111,336 individuals) over and above individual differences in personality[1]. Taking a similar multilevel approach, another set of studies found that implicit bias at the person level showed low temporal stability (test–retest correlation: $r = 0.25$), and was weakly associated with discriminatory behaviour in meta-analyses ($r = 0.14–0.24$), suggesting that implicit bias has relatively low validity as an individual-level construct. However, when aggregated at the county or regional level, implicit bias showed much higher stability (test–retest correlation: $r = 0.72$) and was moderately to strongly associated with historical influences of systemic racism (for example, the regional presence of chattel slavery; $r = 0.37–0.64$)[41,42]. The conclusions afforded by these analyses would have been impossible to draw without a big *N* approach.

Given that experiments cannot feasibly be carried out to examine the causal influence of macro-level factors (such as ecology or culture) on individuals' psychology, big-data researchers have proposed alternative statistical tests to infer causality from archival data, such as the instrumental variable approach used regularly in economics[43]. An ideal instrument is one that cannot theoretically be influenced by a dependent variable of interest, and could exert a mediated effect on a dependent variable only through a focal independent variable. For example, in a study testing the large-scale effects of rice (versus wheat) farming practices on collectivist behaviour, researchers used environmental suitability for rice growing as an instrumental variable. Rice suitability robustly predicted collectivism, presumably only because it facilitated the conditions for rice farming practices. As rice suitability is directly determined by nature (for example, geography or temperature) and could not have been influenced by pre-existing cultural practices, this approach enabled the researchers to rule out reverse causality (that is, that collectivism led to rice suitability)[44]. Similarly, other research used the instrumental variable approach to test whether the number of slaves taken from an ethnic group during the historical slave trade in Africa influenced self-reported trust among their descendants. This study used the geographical distance of ethnic groups from the coast as an instrumental variable, because groups located near the coast were more likely to be forced into the slave trade. At the same time, distance from the coast is, theoretically, uncorrelated with other factors that might influence interpersonal trust. Supporting predictions, individuals from ethnic groups that historically resided near the coast reported less trust in others, an effect the researchers assumed could only have been driven by slave trade activity in Africa[45].

Large-scale datasets of digital records also enable researchers to examine real-world behaviours, including those that occur spontaneously in day-to-day life such as credit or debit transactions[1,46] and mobile app use[47]. Other lines of work have applied natural language processing to social media data to study how the linguistic content of social media posts predicts their spread and virality (indicators of social influence). In one example, the degree of moral-emotional content in tweets was found to be associated with the number of retweets they received, particularly within social networks that share a political ideology[22]. Another study connected more than 1.6 million Reddit users who referenced opioid drugs in their posts to the US states they lived in (based on users' visible geographic tags on Reddit). Across almost a decade, the frequency of synthetic opioid mentions in Reddit posts were positively correlated with the number of documented regional overdose deaths and the number of regional opioid use reports submitted to forensic laboratories. Furthermore, lagged forecasting analyses showed that real-time Reddit mentions of opioids improved prediction of overdose death rates beyond models that relied only on overdose mortality data from 6 months earlier (14.29% reduction in absolute prediction error). These results highlight the potential utility of social media for monitoring trends in problematic drug use in real time[48].

Research using big data has also had a major impact on understanding the consumption of modern digital technologies[49]. Tracking the behaviours of individual website visitors or account holders overcomes shortcomings with traditional self-report approaches, namely that participants can be highly inaccurate in self-reporting activities such as gambling spend[50] or smartphone use[51]. For example, in research on gambling, access to secondary datasets from commercial online platforms enables researchers to view bet-by-bet activity in individual customers[52]. Using this approach, researchers found that loss chasing (the behavioural tendency to continue gambling after a financial loss to recoup the lost money) — often regarded as a hallmark of disordered gambling — can be expressed in multiple ways, including a shorter interval between successive website visits and an increase in wager size following losses[53–55].

An ultimate objective in much of this behavioural big-data research is to design predictive algorithms to detect and intervene in cases of high-risk behaviour. Supervised machine learning, which involves training a statistical model to classify cases into groups (for example, patients versus control individuals) using a wide range of (psychological) input variables, is ideally suited for this task[56]. Random forest models have been found to predict problematic gambling from tracked behavioural data with moderate performance (66–78% accuracy)[57,58]. Similar supervised models have been leveraged in clinical settings to predict responses to depression medication[59] or treatment retention[60]. Approaches that integrate natural language processing and machine learning models trained on social media text have been able to classify people who were previously diagnosed with depression[61–63], suicidal ideation[64] and substance addictions[62,65]. Although these models are not perfectly accurate and performance might vary across samples[66,67], they can increase the predictive power of models that previously relied on self-reported indicators alone[68].

In sum, big data introduces several welcome additions to psychologists' methodological toolkit. Leveraging the statistical power afforded by big data, researchers can examine the complex interplay of effects at multiple levels of analysis, study the behavioural consequences of psychological theories at scale, and make precise predictions about real-world phenomena.

Broader opportunities for psychological science

Psychology has had a long tradition of deductive hypothesis testing: researchers develop theories in a top-down manner, then test the veracity of specific predictions using controlled experiments. This classic 'theories-guide-data' approach has framed the kinds of questions psychologists ask, which usually entail variations of how does X influence Y? Does the effect of X on Y vary according to Z? However, explanatory conclusions drawn from underpowered, non-representative samples are often not replicable, leading to predictions in other domains, cultural contexts or timescales that fail to hold up[69,70]. For example, during the onset of the COVID-19 pandemic, behavioural scientists were quick to propose policies derived from hypotheses grounded in social psychology[71]. However, evidence is mixed on whether predictions made by psychologists about the pandemic were generally accurate. Although one study[72] found support for most of the specific claims in the policy proposals suggested in ref. 71, others found that most expert predictions regarding the pandemic were no more accurate than chance or predictions made by the lay public[73,74]. For scholars wishing to develop theories that have practical implications for addressing pressing, real-world issues, the mixed evidence for forecasting accuracy poses a challenge to the credibility of psychological science.

To address this issue, psychology could leverage big data to answer questions about how temporal, cross-cultural and contextual factors predict real-world psychological trends. For example, applying machine learning to large datasets might enable researchers to answer multifaceted questions about how behaviour Y is likely to change over time and across contexts, the factors that might give rise to changes in Y, and which of these factors matter most. Indeed, psychologists who made predictions about behaviour during the pandemic based on prior data tended to be more accurate than psychologists who made predictions based on theory alone[74]. Thus, big data can help psychology to develop realistic, data-driven theories instead of rehashing old theories that might not be robust to cultural differences across space and time.

In a related vein, psychology's reliance on laboratory and online experiments has promoted a tendency towards overemphasizing internal validity at the expense of external and ecological validity[75,76]. Consequently, many experimental findings might be replicable only in specific, artificial contexts, leading to fragmented and sometimes contradictory theories about the human mind[77]. Indeed, psychologists often develop theories post hoc on the basis of experimental results, and these theories, in turn, influence the design of subsequent experiments to test them (mutual-internal validity). Through this circular process, experiments and the phenomena they purport to capture can become increasingly divorced from reality[70]. This issue is compounded by the fact that many foundational experiments were conducted with non-representative samples in western countries[78].

Although not a panacea to these concerns, a shift towards big data might encourage psychologists to think more deeply about how to develop their theories from real-world rather than laboratory-based observations. With large, representative datasets, psychologists can elevate theories that have predictive power and deprioritize those that do not. Using rich real-world data on lexical associations online (for example, word–object associations in movie or restaurant reviews), cognitive scientists can train algorithmic models based on theories about how people form associations or make decisions. Leveraging emergent patterns in these large-scale datasets, researchers could fine-tune theoretical models to reflect behaviour in the real world. Moreover, if one model demonstrates greater predictive accuracy than others in new samples featuring similar outcomes[79], a strong case could be made for that model's generalizability. By sharpening a focus on real-world phenomena, big data might also encourage researchers to embrace complementary research methods that emphasize people's lived experiences, such as field studies with under-represented populations[80] and archival analyses that document changes in behaviour over time[43].

Ultimately, by facilitating more accurate inferences about future and real-world trends, the increasing use of big data might prioritize predictive power in psychology. Prediction is not about simply finding evidence in support of a model within a given sample. Rather, it refers to the subsequent capacity of that model to make accurate predictions in new samples and contexts[8,81,82]. Importantly, explanatory models are needed in psychology to elucidate causal processes underlying broad theories, but it is often unclear whether these processes generalize beyond the sample studied. Many theoretical models fail to sufficiently predict outcomes beyond the specific context in which they were developed. An explanatory model that is validated in only one experimental context (for example, in a laboratory with undergraduates) and has low predictive power in other comparable contexts should be scrutinized and potentially revised. However, it is important that the right analytic methods are used to ensure predictive validity with big data. Specifically, to ensure the validity of behavioural models trained on large-scale datasets, their accuracy should be examined not only with 'hold-out' test sets (parts of the dataset not used for training), but also with new samples in which similar outcomes are expected.

In sum, big data's affordances of external and ecological validity, in tandem with advances in machine learning, have the potential to shift research priorities in psychological science for the better, enabling the development of robust theories that hold up across contexts.

Risks of big data

Despite its potential to elevate standards for psychology both methodologically and in the questions it pushes researchers to pursue, an increased reliance on big data might also have limitations. Specifically, relying on these datasets could reintroduce longstanding methodological

problems and create new issues that come packaged with the advanced computational approaches that big data typically requires.

Methodological limitations

Feasibility constraints tend to preclude formal experimental designs with big *N* methods. Consequently, most big-data research is correlational, which limits causal inferences. Longitudinal methods that track changes in psychological variables over time have been leveraged to draw inferences that are closer to causal. However, repeated-measures data collection can be costly, and there is a risk of participant attrition. Many psychologists therefore make temporal inferences based on mean-level differences between cohorts[83,84] rather than observations of how individuals actually change over time. However, it is impossible to fully control factors that might confound temporal inferences in multiwave cross-sectional data, such as whether cohorts differ on meaningful traits (for example, identifying as liberal versus conservative) beyond differences in time or age.

To overcome this issue, researchers in fields such as health psychology[85,86] and historical psychology[87,88] have leveraged large-scale longitudinal datasets that enable researchers to estimate lagged relationships between constructs. For example, historical psychologists often sample large corpora of text data published across time as a proxy for psychological change, and some have used regression-based methods to carefully partition cohort differences from change effects[89,90]. However, even with highly granular big *T* datasets, causality can be inferred only if very strict statistical assumptions are met[91]. In many cases, researchers are unable to control for all hidden variables that could influence variables of interest across time.

As discussed above, big-data researchers might draw causal inferences from observational and cross-sectional data using alternative methods such as the instrumental variable approach. Despite the advantages of these statistical techniques, they are not foolproof. Instrumental variable analyses require researchers to select an effective instrument that could have influenced a dependent variable only through an independent variable of interest. This requirement is difficult to meet in psychology and typically requires distal ecological variables (for example, climate factors that are unlikely to directly shape the dependent variable of interest; see ref. 44 for an example). Given the complexity of the human mind and the fact that many dispositions are interrelated and can influence one another in hidden ways, instrumental variable approaches are not feasible for addressing questions that involve proximal processes (for example, how emotions influence goals or how purpose in life influences well-being).

Another problem is that it is not always clear how well measurable or already-measured big-data variables map on to psychological constructs of interest. As researchers lack control over which data are collected in secondary or archival datasets, they often end up using behavioural indicators as indirect proxies of a construct[7]. In the absence of face-valid constructs, theoretically motivated assumptions must be made to determine whether a behavioural outcome (such as retweets on social media) adequately represents a particular construct (such as actual endorsement of the tweet's message). If the goal is to predict real world behaviour, outcomes such as social media activity are viable indicators given their consequences for public discourse and polarization in the modern era. However, to understand what these variables represent from a psychological standpoint, observational inferences should be supplemented with studies in which researchers design measures that capture their theoretical construct of interest. Indeed, experimental findings suggest that retweets might not always represent explicit endorsement of a message and sometimes reflect the opposite[92].

Furthermore, cognitive task indicators such as response latencies and mouse clicks are prone to high error rates, which might hinder meaningful interpretation. For example, response latencies following different attribute–valence pairs in the implicit association test[93] might reflect automatic semantic associations, but could also reflect errors due to random key presses and browser differences, leading to low reliability at the individual level[94]. Although some of these issues are not unique to big-data research, the settings that typically accompany big-data research can exacerbate them. For example, hardware differences such as screen size and output modality (keyboard, mouse control or touchscreen) can significantly influence task performance in cognitive tasks administered through websites[95].

Many publicly available secondary big datasets (such as the World Values Survey or the European Social Survey) expand the scope of psychological research by including a variety of self-report measures collected across multiple waves and countries. However, when designing large-scale, multinational surveys for public use, researchers are often incentivized to include as many measures or items as possible to increase their dataset's utility for addressing a broad range of questions. As these surveys tend to prioritize quantity, the measures used are not always taken from well-validated self-report scales and, in some cases, have been arbitrarily truncated. For example, abridged measures of the Big Five personality traits in the World Values Survey have been critiqued for their poor psychometric properties[96], which makes results derived from these data difficult to trust and compare with research that used established full-length scales.

Indeed, the degree of flexibility offered by big data might be viewed as a double-edged sword: the same datasets that enable rich inferences also enable *P* hacking and the kind of 'fishing expeditions' that contributed to the replicability crisis[97]. In health psychology, exploratory big *V* studies often search for effects of selected variables on multiple outcomes simultaneously[98,99] and emphasize the effects of one or a few variables on a range of outcomes, rather than examining each outcome independently. For researchers with a theory-driven interest in specific dependent variables, this approach can increase family-wise type I error rates (false positives)[34,100]. Moreover, it is easy to find significant correlations below the conventional threshold of $P < 0.05$ in large samples[101]. More complex, robust statistical approaches (such as specification curve analyses, which test how different model specifications might lead to changes in effect sizes[102]) should be applied to large datasets to differentiate spurious from meaningful associations[8]. Ultimately, assessment of specific relationships among variables must be grounded in theory to avoid overinterpreting chance correlations[43].

Machine learning can also be used to computationally identify which emergent trends in big data are meaningful. However, overfitting can produce models that fail to generalize to novel datasets[31]. One study on suicidal ideation claimed to have discovered a neural biomarker with 91% classification accuracy, but other research found that this estimate was inflated because of overfitting. In other words, the model had capitalized on specific idiosyncrasies in the training set that limited its generalizability to new samples[101]. The original article was subsequently retracted. Thus, although big data affords high statistical power, it does not always lead to robustness and replicability. Without strong theoretical foundations and an understanding of best practices in computational psychology, researchers' analytic choices can still drastically bias conclusions[103].

Importantly, these limitations are not necessarily inherent to big data. Instead, they reflect common ways in which big-data tools can be misused, and these issues are more likely to arise with big data than traditional experiments. Following the replication crisis, guardrails were put in place to ensure that laboratory experiments are rigorous, adequately powered and carefully designed[97]. As a result of these new norms that promote robustness, leading psychology journals (such as *Psychological Science*[104]) are less likely to accept manuscripts that interpret potentially spurious correlations derived from underpowered laboratory studies. Similar guardrails are needed for researchers leveraging big data. One promising approach might be to encourage researchers to conduct exploratory analyses in one subset of a large dataset, then preregister and test focal models in the remaining data to minimize analytic over-flexibility. As mentioned above, another approach is to conduct specification curve analyses to assess whether significant associations hold across the vast array of analytic choices that are typically available with big data and rule out spurious correlations[105]. In cases where behavioural indicators are used as proxies for psychological constructs, journals could require that researchers provide data that confirm these indicators' validity. Despite the promise of these potential guidelines, they are currently either underspecified or non-normative in psychology. It is therefore important to remind scholars of the risks that could arise with any study but have been less thoroughly considered for studies that use big data.

Broader risks for psychological science

The limitations described above point to a lack of theoretical maturity underlying the interpretation of big datasets, which potentially opens the field to the risk of a new 'theory crisis'. Specifically, big data offer predictive power, but many of these approaches lack explanatory power. Thus, although big data have the potential to shift psychology's focus towards increasingly accurate predictions about real-world, future events, they might simultaneously undervalue understanding the 'why' behind psychological effects.

Scholars have argued that that are too many effects in psychology and not enough models to properly explain them[106]. By virtue of enabling even tiny correlations to be statistically significant using standard inferential tests, big data might introduce more unexplained effects into this pool. Without a thoughtful and thorough consideration of context and practical relevance[31], a grab-bag of statistically significant associations might increase theoretical unknowns. That is, without an understanding of why these associations exist, it is difficult to distinguish meaningful findings from statistical artefacts, an issue that compounds existing challenges to determining which associations hold across contexts and timescales. Big data should be accompanied — and informed — by 'big theory' to avoid fragmented theory development[77].

As researchers often have little control over which questions or measures are included in many large datasets, big-data studies often lack the intentional design needed to answer specific questions that can test explanatory models[8,79]. This structural impediment, along with the pressure to report robust results with big *N*s, might eventually select for big-data research questions based on a limited set of variables that already exist in public self-report data (such as life satisfaction, well-being and socioeconomic status). In effect, psychology could see a cultural shift from 'theories-guide-data' to 'data-guide-theories', which would limit the potential for new, disruptive ideas to flourish.

To be clear, data-guided theorizing is not necessarily a problem in and of itself, and it can help to refine theory by accounting for real-world phenomena. However, we caution against an over-reliance on data — no matter how big and externally valid — as a starting point for theory development. A complex phenomenon in the real world can often be explained by multiple theories, and identifying the correct one requires strong, meta-theoretical assumptions[105] that might be incorrect. Instead of developing theory from associations found in big data alone, psychologists should first formalize theories with carefully delineated parameters[106–109]. In doing so, researchers could focus on the theories most likely to have explanatory and predictive power, enabling subsequent inferences from big data to be theoretically informed and generalizable.

Moreover, reliance on big-data approaches could do little to change — and potentially even exacerbate — the publish-or-perish mentality that pressures psychologists to publish quick, low-quality research[110]. The best big-data studies often require advanced statistical techniques, careful robustness tests and the creative use of archival datasets. Committing to these approaches requires time and effort. Researchers might feel that they must choose between developing the specialized skills in computational methods recommended for big data and careful theory building. Indeed, cultural models of psychological science suggest that a focus on methods can crowd out theoretical innovation[111].

It is possible that the increasing use of large language models (for example, ChatGPT or Claude Code) for coding will enable researchers to adopt advanced methods such as machine learning more quickly[112]. However, the standards for big-data research continue to evolve rapidly, and the statistical training psychologists receive in graduate programmes will probably struggle to keep pace with these developments[113,114]. The pressure to publish quickly might steer researchers away from learning advanced big-data methods and encourage them to tackle low-hanging fruit using more basic cross-sectional analyses that are prone to spurious effects.

Of course, the publish-or-perish incentive structure in academia influences all forms of research, not only those that involve big data. However, in the wake of the replication crisis, experimentalists now tend to strive for greater power and rigour than before, and, on average, data collection for experiments is slower and more costly than for studies using large-scale secondary or archival datasets. Importantly, whereas principal investigators might be hesitant to devote limited funds to a time-consuming and costly series of well-powered programmatic laboratory or online experiments, the barrier to entry for free open-source datasets is low or even non-existent. Consequently, although a publish-or-perish incentive structure can also motivate 'quick and dirty' experiments, it has become more difficult to quickly publish those kinds of laboratory studies compared with big-data projects. Unique affordances of big data (scale, flexibility and availability) might interact with publish-or-perish incentives in ways that uniquely promote problematic practices that have become less tenable with traditional methods.

In sum, the limitations of big data combined with the incentive structure of psychological science could promote the natural selection of bad big-data science[115]. In the future, instead of underpowered, *P*-hacked experiments, psychology might see the proliferation of high-powered but theory-bereft observational research that does little to stimulate new explanatory models. This outcome might emerge hand in hand with an over-reliance on a few powerful but limited big-data sources, such as the World Values Survey and the Gosling–Potter Internet Personality project. In short, big-data methods can be used to advance theory when done correctly, but researchers must be incentivized to do so.

Moving forward

An important question for the field moving forward is how to harness the full potential of big data without succumbing to its pitfalls. Importantly, the predictive power of big data and the explanatory power of experiments need not exist in siloes and can be harmonized in projects that answer focused questions. Here, we describe three ways in which big data and experiments can work in tandem, offering the best of both worlds and ultimately promoting theoretically rigorous big-data science (Fig. 1): conducting big *N* experiments, evaluating computational models trained on big *N* data using smaller *N* experiments, and providing convergent evidence for a multifaceted theory with both big *N* correlational studies and smaller *N* experiments.

Running big *N* experiments

First, researchers can use online platforms to conduct large-scale experiments, which directly sidesteps the limitations of the big observational datasets that we have focused on here. Indeed, some scholars have leveraged partnerships with companies that collect social media data, or launched publicly accessible websites that have gone viral, to conduct randomized experiments in ecologically valid contexts[50,116–118]. For example, one set of researchers partnered with digital advertisers to examine how personality traits shape engagement with persuasive messages[16]. As the personality traits of millions of Facebook users can be reliably inferred from their online behaviour (such as 'likes' and status updates)[119], the researchers were able to track how people with high versus low levels of extraversion and openness-to-experience engaged with ads that either did or did not match their personality traits. The results provided compelling evidence that psychological targeting based on personality can increase engagement with ads and influence real-world purchasing behaviour[16]. On a theoretical level, these findings offered large-scale, ecologically valid evidence for the causal role of personality–message fit in determining the persuasiveness of social messages.

Other researchers have conducted big *N* experiments without partnering with social media companies or digital advertisers. For example, a team of interdisciplinary scholars developed their own online experimental platform called The Moral Machine to understand differences in how people evaluate moral dilemmas related to autonomous driving cars[120]. Visitors to their website indicated their anticipated responses to a variety of scenarios, including the trolley problem (a classic sacrificial dilemma in which participants are asked whether they would divert an incoming trolley so that it kills one person instead of multiple people). Researchers compared participants' preferences for utilitarian-decision scenarios that varied in key features (such as number and demographics of potential casualties). Results from the platform led to novel findings about moral decision-making in high-stakes contexts[120–122].

Applying big-data-derived computational models to smaller-scale experiments

Second, researchers can first train models on large-scale data that capture emergent patterns in how people make choices and decisions (for example, reading lists, shopping preferences or online consumption data) and then test the validity of these computational models in predicting decision-making in smaller-scale experiments[123–126]. For example, in one study researchers theorized that algorithm recommendations are often based on the preferences of others (for example, Spotify algorithms base song recommendations on the average preferences of similar users), whereas humans tend to make recommendations using a Bayesian generalization strategy that involves assessing how items generally 'fit' together (for example, friends recommend songs on the basis of genre fit)[124]. To test their hypothesis, the researchers first trained algorithmic recommendation models to incorporate these different strategies using large-scale datasets of music playlists and reading lists found online (N = ~150,000). They next conducted experiments (N = 400) in which participants judged the quality of recommendations made by these models and generated their own suggestions for items that should be added to partially completed playlists and reading lists. Crucially, models using the Bayesian generalization strategy (which mirrors how humans typically make recommendations) produced recommendations that participants rated as having the highest quality, and these recommendations were most strongly correlated with participants' own list suggestions[124].

In another study, researchers conducted a large-scale competition in which cognitive scientists were invited to submit computational models that would best predict risky choice behaviour[126]. Some models were purely data-driven (trained exclusively in a bottom-up fashion on large-scale experimental datasets), whereas others included theory-driven machine learning parameters. A hybrid model that

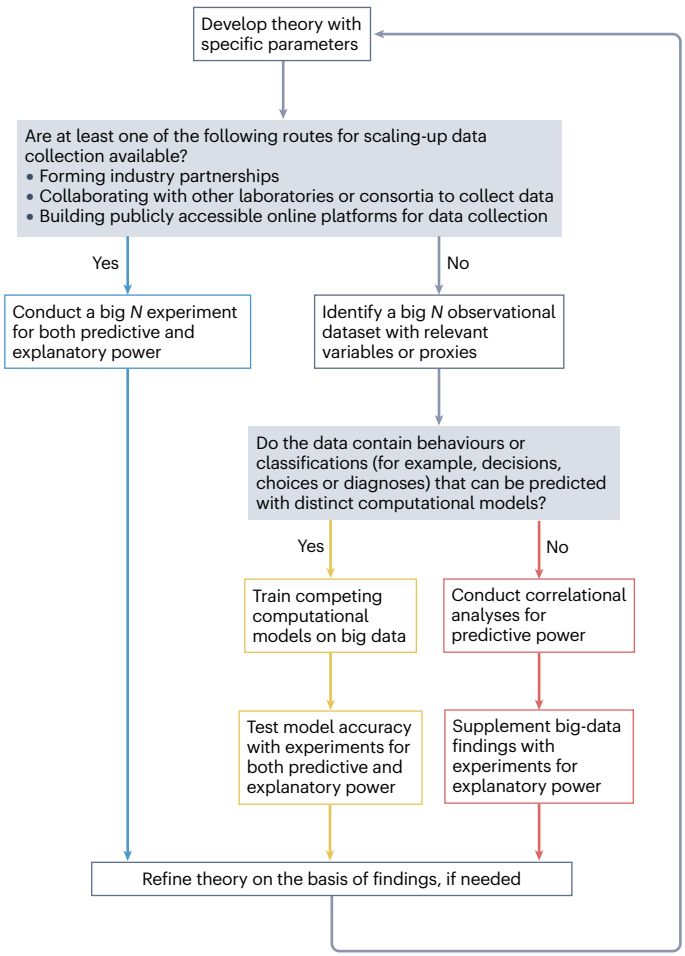

Fig. 1 | Theory-integrated pipelines for big-data research in psychology. Three possible, non-exhaustive and non-mutually-exclusive pipelines for integrating big data with theory development. The grey boxes are critical questions researchers will need to ask at each stage to decide which pipeline is best for their research.

incorporated both approaches outperformed competing models in predicting risky choice behaviours. Importantly, a purely data-driven version of the model performed worse than a purely theory-driven version of the model, supporting arguments[107] about the benefits of developing formalized computational theories before data collection[126]. Along with prior work by the same research team[125], these results suggest that theory development based on data alone might be insufficient for making accurate forecasts about the real world[87]. Importantly, this experiment, as well as other work that adopted a similar two-step approach[125,127], had the same set of participants complete multiple tasks. This big *V* set up within an experimental design enables researchers to test the generalizability of models across tasks that involve similar choice paradigms, providing the opportunity for cleaner tests of predictive power.

Combining big correlational data and experiments

Finally, big-data studies can be combined with smaller-scale experiments to provide complementary evidence for a theory. This approach contributes to a cumulative science in which each study builds on and improves the other and might be more feasible than the prior two approaches. Psychologists often lack the resources to partner with companies or to launch websites that can reach a large audience, and computational modelling can be difficult to learn.

Some researchers have explored 'scaling up' from experiments to big data: running controlled laboratory or online experiments first, then complementing their conclusions with big field data or cross-sectional studies in the real world[128,129]. For example, in one study the researchers used online experiments ($N = 1{,}562$) and secondary big *N* data on personality ($N = 206{,}587$) to address the question of which comes first: stereotypes or real-world inequalities[129]. Specifically, they examined how discrimination in contemporary Chinese society is shaped by astrological stereotypes, which are relatively novel cultural constructs in China and should therefore not be based on pre-existing social inequalities. In small-scale experiments, Chinese participants discriminated against individuals with specific astrological signs. To rule out the possibility that these stereotypes were based on some kind of reality, the researchers used a big *N* dataset ($N = 173{,}709$) and archival employee data from a large Chinese conglomerate ($N = 32{,}878$) to show that astrological signs are not associated with personality traits or job performance[129].

Other approaches involve 'scaling down': beginning with big data to identify robust, cross-sectional trends or differences, then conducting experiments to test whether these observed differences replicate[130–133]. For example, one study used this approach to examine the causal association between regional pathogen prevalence and a tendency to value moral purity[130]. The researchers found a robust cross-sectional association between pathogen prevalence and concern for purity across large-scale datasets that included region-level ($N = 2{,}834$ US counties) and country-level ($N = 67$ countries) indicators. Next, to test the linguistic generalizability of this link, they examined word embeddings (a natural language processing method) in a corpus of texts written in different languages and found that words denoting pathogens strongly co-occurred with words denoting purity concerns. To complement these big-data approaches, they ran a laboratory study ($N = 320$) showing that exposure to images of pathogens led to stronger purity concerns than exposure to non-pathogen-related images[130].

Using natural language processing approaches with big data to inform subsequent experimental design can also enhance the study

Table 1 | Summary of opportunities and risks of big data and caveats to these points

Opportunity or risk	Feature or implication	Explanation	Caveats
Methodological features			
Opportunity	Statistical power	Big data allow for enough statistical power to identify small effects	Effects might be significant but not theoretically meaningful Advanced robustness analyses (for example, specification curve analyses) are needed to rule out spurious associations
	Behavioural evidence	Large-scale archival datasets go beyond self-report to track naturalistic behaviours	Researchers often cannot choose which behaviours were measured and must therefore use proxies for intended constructs
	Multilevel modelling	Big data enable examination of individual-level versus group-level effects and how they might interact	Advanced statistical knowledge of multilevel modelling techniques might be required to appropriately account for clustering Not all observational big datasets are structured at multiple levels
Risk	Constrained data structure	Pre-collected datasets might be limited in construct validity and causal inference	Real-world behaviours are still important and consequential even if unexplained Creative statistical techniques can approximate causal inference, albeit imperfectly
	Analytic flexibility	The novelty of big-data science approaches can lead to biases in analytic decisions	Performance of different models can be compared in held-out test sets or new samples Analytic plans can be preregistered to minimize biases
Broader implications			
Opportunity	Predicting more	Psychologists might increasingly value theories that speak to real-world behaviour and enable predictions about future trends	It is difficult to know whether generalizability will be sustained without identification of causal mechanisms Real-world observations can be explained by multiple theories
Risk	Explaining less	Big data might exacerbate existing problems of fragmented theoretical development in psychology	Big data studies can be supplemented by experiments Real-world data can be used to refine and/or eliminate existing theories The theoretical rigour of big-data approaches is moderated by publish-or-perish incentives

of more proximal cognitive processes. For example, in one study researchers applied semantic similarity analyses to a corpus of around 4,000 nineteenth-century English books (~240 million words), and found that words sharing phonesthemes (sounds attached to words) such as 'gl-' in light-related words such as 'glimmer', 'gleam' and 'glow', tend to be more semantically similar than random words that lack these sounds. Next, using the phonesthemes identified from this corpus, the researchers conducted a laboratory experiment with repeated measures to show that participants ($N = 19$) automatically associate these sounds with meaning, even for made-up words that do not exist in the English language (for example, 'glybe' is automatically assumed to have a meaning related to light, vision or shine)[131]. This multimethod package provided strong evidence that sounds influence listeners' assumptions about the meanings of unfamiliar words.

In addition to publishing big-data-powered and experimental studies in a single paper, as in the examples above, research teams might achieve similar aims by publishing a series of papers that test the same theory or collaborating with other teams interested in the same topic. Some laboratories might specialize in advanced, computational big-data approaches or have proprietary access to large-scale datasets, whereas others might have experience of conducting in-person experiments. Working together, they might collectively contribute to understanding multiple components of a theory without one research team being pressured to single-handedly take on the burden of methodological diversity.

Conclusion

In this Perspective, we aimed to paint a balanced portrait of the implications of the rise of big data for psychological science. On the one hand, big data enable researchers to address questions that have eluded the discipline, shifting psychology's emphasis towards understanding how people behave in the real world and in the future. On the other hand, because of the persistence of a fast-paced, competitive academic system that rewards less ambitious approaches to data collection, big data might do little to change — and could even worsen — psychology's current issues with fragmented theory development.

Taking both the opportunities and risks into account (Table 1), we argue that big data need not replace psychology's traditional methodological paradigms, and that the best research packages are those that find creative ways to implement both. As prior scholars have argued, methodological variety is essential to advancing theory[134], and we have highlighted examples across diverse areas of psychological science that strategically used a mixture of methods to address multiple facets of a theory. Among these, several stand out as one-off big N experiments that successfully achieve both explanatory and predictive power. Nonetheless, the barrier to entry for conducting these studies is not trivial; they often require partnering with organizations with profit-motivated (rather than knowledge-motivated) interests, as well as mastering state-of-the-art computational techniques.

We therefore recommend that psychologists combine large-scale observational data from public sources with smaller N experiments to provide convergent evidence for a theoretical model. Big data sources need not be limited to self-report surveys that have reached a large audience online, and we encourage researchers to think creatively when identifying archival datasets (for example, lottery ticket sales and outcomes of local sporting events[135], regional weather data and protest activity[136]) that might align with their research questions. Each study alone is likely to provide a specific but insufficient lens into a theory; laboratory experiments can elucidate causality but are often unable to capture behavioural patterns in the real world, whereas cross-sectional big N studies can identify real-world trends but often fail to explain them. Taken together, each component addresses the limitations of the other and enriches theory development.

Moving beyond these quantitative approaches, researchers might also consider qualitative approaches to psychology, which offer interpretive power in ways that quantitative big data (predictive power) and experiments (explanatory power) cannot. Some researchers have called for a more subjectivist lens to understanding lived experiences with greater detail and nuance[137]. Although natural language processing methods involve using computational models to quantitatively examine trends in qualitative data[20,23,24,87], psychology theories could also be enriched by highly structured and rigorous qualitative research[138,139] that dives deep into singular subjective experiences instead of aggregating across large-scale trends. Inferences made from qualitative research could inform the subsequent design of experiments and big-data studies. Predictive power, explanatory power and interpretive power can collectively constitute a holistic, robust science.

More broadly, moving beyond mixed methods, we encourage psychology departments to actively incentivize slow science[140], which is essential for nurturing creative and robust research[115]. Part of this systemic change means rewarding job candidates who have published fewer but higher-quality papers that incorporate diverse methods, as well as ambitious theoretically driven work, even if it does not always provide completely 'clean' significant findings. The reality of human psychology is likely to be far messier than what many published psychology papers suggest, and contradictions that emerge between inferences from big data and experiments need not debunk entire models. Instead, this messiness can invigorate psychologists to reflect on the nuances of existing theories and develop new ones that might expand our understanding of the human mind.

Published online: 03 August 2026

References

1. Ebert, T., Götz, F. M., Gladstone, J. J., Müller, S. R. & Matz, S. C. Spending reflects not only who we are but also who we are around: the joint effects of individual and geographic personality on consumption. *J. Pers. Soc. Psychol.* **121**, 378–393 (2021).
2. Joel, S. et al. Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proc. Natl Acad. Sci. USA* **117**, 19061–19071 (2020).
3. Stachl, C. et al. Predicting personality from patterns of behavior collected with smartphones. *Proc. Natl Acad. Sci. USA* **117**, 17680–17687 (2020).
4. Harari, G. M. & Gosling, S. D. Understanding behaviours in context using mobile sensing. *Nat. Rev. Psychol.* **2**, 767–779 (2023).
5. Atari, M., Henrich, J. & Schulz, J. The chronospatial revolution in psychology. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-025-02229-y (2025).
6. Varnum, M. E. W., Baumard, N., Atari, M. & Gray, K. Large language models based on historical text could offer informative tools for behavioral science. *Proc. Natl Acad. Sci. USA* **121**, e2407639121 (2024).
7. Adjerid, I. & Kelley, K. Big data in psychology: a framework for research advancement. *Am. Psychol.* **73**, 899–917 (2018).
8. Peters, H., Marrero, Z. & Gosling, S. D. The big data toolkit for psychologists: data sources and methodologies. In *The Psychology of Technology: Social Science Research in the Age of Big Data* (ed. Matz, S. C.) https://doi.org/10.1037/0000290-004 (American Psychological Association, 2022).
9. Gosling, S. D., Vazire, S., Srivastava, S. & John, O. P. Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *Am. Psychol.* **59**, 93–104 (2004).
10. Roemer, L. et al. Beyond age and generations: how considering period effects reshapes our understanding of personality change. *J. Pers. Soc. Psychol.* **130**, 129–159 (2025).
11. Srivastava, S., John, O. P., Gosling, S. D. & Potter, J. Development of personality in early and middle adulthood: set like plaster or persistent change? *J. Pers. Soc. Psychol.* **84**, 1041–1053 (2003).
12. Robins, R. W. et al. A longitudinal study of personality change in young adulthood. *J. Pers.* **69**, 617–640 (2001).
13. Hinds, J. & Joinson, A. Human and computer personality prediction from digital footprints. *Curr. Dir. Psychol. Sci.* **28**, 204–211 (2019).

14. Kosinski, M., Stillwell, D. & Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl Acad. Sci. USA* **110**, 5802–5805 (2013).
15. Youyou, W., Kosinski, M. & Stillwell, D. Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl Acad. Sci. USA* **112**, 1036–1040 (2015).
16. Matz, S. C., Kosinski, M., Nave, G. & Stillwell, D. J. Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl Acad. Sci. USA* **114**, 12714–12719 (2017).
17. Islam, M.dR. et al. Depression detection from social network data using machine learning techniques. *Health Inf. Sci. Syst.* **6**, 8 (2018).
18. Priya, A., Garg, S. & Tigga, N. P. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Comput. Sci.* **167**, 1258–1267 (2020).
19. Feuerriegel, S. et al. Using natural language processing to analyse text data in behavioural science. *Nat. Rev. Psychol.* **4**, 96–111 (2025).
20. Charlesworth, T. E. S., Sanjeev, N., Hatzenbuehler, M. L. & Banaji, M. R. Identifying and predicting stereotype change in large language corpora: 72 groups, 115 years (1900–2015), and four text sources. *J. Pers. Soc. Psychol.* **125**, 969–990 (2023).
21. Rathje, S. et al. GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl Acad. Sci.USA* **121**, e2308950121 (2024).
22. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. & Van Bavel, J. J. Emotion shapes the diffusion of moralized content in social networks. *Proc. Natl Acad. Sci. USA* **114**, 7313–7318 (2017).
23. Gabriel, R. A. et al. On the development and validation of large language model-based classifiers for identifying social determinants of health. *Proc. Natl Acad. Sci. USA* **121**, e2320716121 (2024).
24. Scheffer, M., van de Leemput, I., Weinans, E. & Bollen, J. The rise and fall of rationality in language. *Proc. Natl Acad. Sci. USA* **118**, e2107848118 (2021).
25. Schwartz, H. A. et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE* **8**, e73791 (2013).
26. Lazer, D. et al. Computational social science. *Science* **323**, 721–723 (2009).
27. Qiu, L., Chan, S. H. M. & Chan, D. Big data in social and psychological science: theoretical and methodological issues. *J. Comput. Soc. Sc.* **1**, 59–66 (2018).
28. Favaretto, M., Clercq, E. D., Schneble, C. O. & Elger, B. S. What is your definition of big data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE* **15**, e0228987 (2020).
29. Chen, E. E. & Wojcik, S. P. A practical guide to big data research in psychology. *Psychol. Methods* **21**, 458–474 (2016).
30. Funder, D. C. & Ozer, D. J. Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).
31. Götz, F. M., Gosling, S. D. & Rentfrow, P. J. Effect sizes and what to make of them. *Nat. Hum. Behav.* **8**, 798–800 (2024).
32. Brady, H. E. The challenge of big data and data science. *Annu. Rev. Polit. Sci.* **22**, 297–323 (2019).
33. Cao, L. Data science: a comprehensive overview. *ACM Comput. Surv.* **50**, 1–42 (2018).
34. Cohen, J. Things I have learned (so far). *Am. Psychol.* **45**, 1304–1312 (1990).
35. Yarkoni, T. Big correlations in little studies: inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspect. Psychol. Sci.* **4**, 294–298 (2009).
36. Götz, F. M., Gosling, S. D. & Rentfrow, P. J. Small effects: the indispensable foundation for a cumulative psychological science. *Perspect. Psychol. Sci.* **17**, 205–215 (2022).
37. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
38. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
39. Zhi, D. et al. Triple interactions between the environment, brain, and behavior in children: an ABCD study. *Biol. Psychiatry* **95**, 828–838 (2024).
40. Miller, D. T. & Laurin, K. History of social psychology: four enduring tensions. In *The Handbook of Social Psychology* (eds Gilbert, D. T., Fiske, S. T., Mendes, W. B. & Finkel, E. J.) https://doi.org/10.70400/DCSX1997 (Situational Press, 2025).
41. Payne, B. K. & Hannay, J. W. Implicit bias reflects systemic racism. *Trends Cognit. Sci.* **25**, 927–936 (2021).
42. Vuletich, H. A. & Payne, B. K. Stability and change in implicit bias. *Psychol. Sci.* **30**, 854–862 (2019).
43. Muthukrishna, M., Henrich, J. & Slingerland, E. Psychology as a historical science. *Annu. Rev. Psychol.* **72**, 717–749 (2021).
44. Talhelm, T. et al. Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science* **344**, 603–608 (2014).
45. Nunn, N. & Wantchekon, L. The slave trade and the origins of mistrust in Africa. *Am. Econ. Rev.* **101**, 3221–3252 (2011).
46. Muggleton, N. et al. The association between gambling and financial, social and health outcomes in big financial data. *Nat. Hum. Behav.* **5**, 319–326 (2021).
47. Harari, G. M. et al. Sensing sociability: individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *J. Pers. Soc. Psychol.* **119**, 204–228 (2020).
48. Smith, D. A. et al. Monitoring the opioid epidemic via social media discussions. *npj Digit. Med.* **8**, 284 (2025).
49. Flayelle, M. et al. A taxonomy of technology design features that promote potentially addictive online behaviours. *Nat. Rev. Psychol.* **2**, 136–150 (2023).
50. Heirene, R. M. & Gainsbury, S. M. Encouraging and evaluating limit-setting among on-line gamblers: a naturalistic randomized controlled trial. *Addiction* **116**, 2801–2813 (2021).
51. Ellis, D. A., Davidson, B. I., Shaw, H. & Geyer, K. Do smartphone usage scales predict behavior? *Int. J. Hum. Comput. Stud.* **130**, 86–92 (2019).
52. Deng, X., Lesch, T. & Clark, L. Applying data science to behavioral analysis of online gambling. *Curr. Addict. Rep.* **6**, 159–164 (2019).
53. Banerjee, N., Chen, Z., Clark, L. & Noël, X. Behavioural expressions of loss-chasing in gambling: a systematic scoping review. *Neurosci. Biobehav. Rev.* **153**, 105377 (2023).
54. Edson, T. C., Louderback, E. R., Tom, M. A., McCullock, S. P. & LaPlante, D. A. Exploring a multidimensional concept of loss chasing using online sports betting records. *Int. Gambl. Stud.* **24**, 306–324 (2024).
55. Zhang, K., Rights, J. D., Deng, X., Lesch, T. & Clark, L. Within-session chasing of losses and wins in an online eCasino. *Sci. Rep.* **14**, 20353 (2024).
56. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118 (2018).
57. Finkenwirth, S., MacDonald, K., Deng, X., Lesch, T. & Clark, L. Using machine learning to predict self-exclusion status in online gamblers on the PlayNow.com platform in British Columbia. *Int. Gambl. Stud.* **21**, 220–237 (2021).
58. Murch, W. S. et al. Using machine learning to retrospectively predict self-reported gambling problems in Quebec. *Addiction* **118**, 1569–1578 (2023).
59. Chekroud, A. M. et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* **3**, 243–250 (2016).
60. Curtis, B. et al. AI-based analysis of social media language predicts addiction treatment dropout at 90 days. *Neuropsychopharmacol.* **48**, 1579–1585 (2023).
61. Eichstaedt, J. C. et al. Facebook language predicts depression in medical records. *Proc. Natl Acad. Sci. USA* **115**, 11203–11208 (2018).
62. Liu, T. et al. Detecting symptoms of depression on Reddit. In *Proc. 15th ACM Web Science Conference 2023* https://doi.org/10.1145/3578503.3583621 (Association for Computing Machinery, 2023).
63. Milintsevich, K., Sirts, K. & Dias, G. Towards automatic text-based estimation of depression through symptom prediction. *Brain Inf.* **10**, 4 (2023).
64. Roy, A. et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *npj Digit. Med.* **3**, 78 (2020).
65. Matero, M. et al. Using daily language to understand drinking: multi-level longitudinal differential language analysis. In *Proc. 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)* (eds Yates, A. et al.) 133–144 (Association for Computational Linguistics, 2024).
66. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017).
67. Rai, S. et al. Key language markers of depression on social media depend on race. *Proc. Natl Acad. Sci. USA* **121**, e2319837121 (2024).
68. Whelan, R. et al. Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185–189 (2014).
69. Grossmann, I., Varnum, M. E. W., Hutcherson, C. A. & Mandel, D. R. When expert predictions fail. *Trends Cognit. Sci.* **28**, 113–123 (2024).
70. Lin, H., Werner, K. M. & Inzlicht, M. Promises and perils of experimentation: the mutual-internal-validity problem. *Perspect. Psychol. Sci.* **16**, 854–863 (2021).
71. van Bavel, J. J. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
72. Ruggeri, K. et al. A synthesis of evidence for policy from behavioural science during COVID-19. *Nature* **625**, 134–147 (2024).
73. Hutcherson, C. A. et al. On the accuracy, media representation, and public perception of psychological scientists' judgments of societal change. *Am. Psychol.* **78**, 968–981 (2023).
74. Grossmann, I. et al. Insights into the accuracy of social scientists' forecasts of societal change. *Nat. Hum. Behav.* **7**, 484–501 (2023).
75. Baumeister, R. F., Vohs, K. D. & Funder, D. C. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* **2**, 396–403 (2007).
76. Cialdini, R. B. We have to break up. *Perspect. Psychol. Sci.* **4**, 5–6 (2009).
77. Muthukrishna, M. & Henrich, J. A problem in theory. *Nat. Hum. Behav.* **3**, 221–229 (2019).
78. Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature* **466**, 29–29 (2010).
79. Johns, B. T., Jamieson, R. K. & Jones, M. N. The continued importance of theory: lessons from big data approaches to language and cognition. In *Big Data in Psychological Research* (eds Woo, S. E., Tay, L. & Proctor, R. W.) https://doi.org/10.1037/0000193-013 (American Psychological Association, 2020).
80. Jachimowicz, J. M. Embracing field studies as a tool for learning. *Nat. Rev. Psychol.* **1**, 249–250 (2022).
81. Shmueli, G. To explain or to predict? *Stat. Sci.* **25**, 289–310 (2010).
82. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
83. Charlesworth, T. E. S. & Banaji, M. R. Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychol. Sci.* **30**, 174–192 (2019).
84. Li, L. M. W. & Bond, M. H. Value change: analyzing national change in citizen secularism across four time periods in the world values survey. *Soc. Sci. J.* **47**, 294–306 (2010).
85. Jokela, M., Airaksinen, J., Kivimäki, M. & Hakulinen, C. Is within–individual variation in personality traits associated with changes in health behaviours? Analysis of seven longitudinal cohort studies. *Eur. J. Pers.* **32**, 642–652 (2018).
86. Luo, J. et al. Personality and health: disentangling their between-person and within-person relationship in three longitudinal studies. *J. Pers. Soc. Psychol.* **122**, 493–522 (2022).

87. Jackson, J. C., Gelfand, M., De, S. & Fox, A. The loosening of American culture over 200 years is associated with a creativity–order trade-off. *Nat. Hum. Behav.* **3**, 244–250 (2019).
88. Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P. & Henrich, J. The church, intensive kinship, and global psychological variation. *Science* **366**, eaau5141 (2019).
89. Sobchuk, O. & Beheim, B. Does literature evolve one funeral at a time? *Proc. Biol. Sci.* **292**, 20242033 (2025).
90. Underwood, T., Kiley, K., Shang, W. & Vaisey, S. Cohort succession explains most change in literary culture. *Soc. Sci.* **9**, 184–205 (2022).
91. Mulder, J. D. & Hamaker, E. L. Three extensions of the random intercept cross-lagged panel model. *Struct. Equ. Model.* **28**, 638–648 (2021).
92. Heltzel, G. & Laurin, K. Why Twitter sometimes rewards what most people disapprove of: the case of cross-party political relations. *Psychol. Sci.* **35**, 976–994 (2024).
93. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
94. Gawronski, B. & Hahn, A. Implicit measures: procedures, use, and interpretation. In *Measurement in Social Psychology* (eds. Blanton, H., Jessica, M. L. & Gregory, D. W.) 29–55 (Routledge, 2018).
95. Passell, E. et al. Cognitive test scores vary with choice of personal digital device. *Behav. Res. Methods* **53**, 2544–2557 (2021).
96. Ludeke, S. G. & Larsen, E. G. Problems with the Big Five assessment in the World Values Survey. *Pers. Individ. Differ.* **112**, 103–105 (2017).
97. Nosek, B. A. et al. Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, 719–748 (2022).
98. Chen, Y., Harris, S. K., Worthington, E. L. Jr. & VanderWeele, T. J. Religiously or spiritually-motivated forgiveness and subsequent health and well-being among young adults: an outcome-wide analysis. *J. Posit. Psychol.* **14**, 649–658 (2019).
99. Long, K. N. G., Wilkinson, R., Cowden, R. G., Chen, Y. & VanderWeele, T. J. Hope in adolescence and subsequent health and well-being in adulthood: an outcome-wide longitudinal study. *Soc. Sci. Med.* **347**, 116704 (2024).
100. García-Pérez, M. A. Use and misuse of corrections for multiple testing. *Methods Psychol.* **8**, 100120 (2023).
101. Orben, A. & Lakens, D. Crud (re)defined. *Adv. Methods Pract. Psychol. Sci.* **3**, 238–247 (2020).
102. Nelson, L. D., Simmons, J. & Simonsohn, U. Psychology's renaissance. *Annu. Rev. Psychol.* **69**, 511–534 (2018).
103. Coles, N. A. et al. Big team science reveals promises and limitations of machine learning efforts to model physiological markers of affective experience. *R. Soc. Open Sci.* **12**, 24 (1778).
104. Vazire, S. The next chapter for *Psychological Science*. *Psychol. Sci.* **35**, 703–707 (2024).
105. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve analysis. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).
106. van Rooij, I. & Baggio, G. Theory before the test: how to build high-verisimilitude explanatory theories in psychological science. *Perspect. Psychol. Sci.* **16**, 682–697 (2021).
107. Guest, O. & Martin, A. E. How computational modeling can force theory building in psychological science. *Perspect. Psychol. Sci.* **16**, 789–802 (2021).
108. Gray, K. How to map theory: reliable methods are fruitless without rigorous theory. *Perspect. Psychol. Sci.* **12**, 731–741 (2017).
109. Syed, M. Elegant theories and the problems of social psychology. *Eur. Rev. Soc. Psychol.* **36**, 298–325 (2025).
110. Brembs, B. et al. Replacing academic journals. *R. Soc. Open Sci.* **10**, 230206 (2023).
111. Gervais, W. M. Practical methodological reform needs good theory. *Perspect. Psychol. Sci.* **16**, 827–843 (2021).
112. Hoogeveen, S. et al. A many-analysts approach to the relation between religiosity and well-being. *Relig. Brain Behav.* **13**, 237–283 (2023).
113. Howard, A. L. Graduate students need more quantitative methods support. *Nat. Rev. Psychol.* **3**, 140–141 (2024).
114. Vezzoli, M. & Zogmaister, C. An introductory guide for conducting psychological research with big data. *Psychol. Methods* **28**, 580–599 (2023).
115. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* **3**, 160384 (2016).
116. Bapna, R., Ramaprasad, J., Shmueli, G. & Umyarov, A. One-way mirrors in online dating: a randomized field experiment. *Manag. Sci.* **62**, 3100–3122 (2016).
117. Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl Acad. Sci. USA* **111**, 8788–8790 (2014).
118. Muchnik, L., Aral, S. & Taylor, S. J. Social influence bias: a randomized experiment. *Science* **341**, 647–651 (2013).
119. Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V. & Stillwell, D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol.* **70**, 543–556 (2015).
120. Awad, E. et al. The moral machine experiment. *Nature* **563**, 59–64 (2018).
121. Awad, E., Dsouza, S., Bonnefon, J.-F., Shariff, A. & Rahwan, I. Crowdsourcing moral machines. *Commun. ACM.* **63**, 48–55 (2020).
122. Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl Acad. Sci. USA* **117**, 2332–2337 (2020).
123. Bhatia, S. & Stewart, N. Naturalistic multiattribute choice. *Cognition* **179**, 71–88 (2018).
124. Bourgin, D. D., Abbott, J. T. & Griffiths, T. L. Recommendation as generalization: using big data to evaluate cognitive models. *J. Exp. Psychol. Gen.* **150**, 1398–1409 (2021).
125. Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. & Griffiths, T. L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**, 1209–1214 (2021).
126. Plonsky, O. et al. Predicting human decisions with behavioural theories and machine learning. *Nat. Hum. Behav.* **9**, 2271–2284 (2025).
127. Zhu, J.-Q., Peterson, J. C., Enke, B. & Griffiths, T. L. Capturing the complexity of human strategic decision-making with machine learning. *Nat. Hum. Behav.* **9**, 2114–2120 (2025).
128. Jordan, J. J. & Sommers, R. Sexual assault victims face a penalty for adjacent consent. *Proc. Natl Acad. Sci. USA* **121**, e2403609121 (2024).
129. Lu, J. G., Liu, X. L., Liao, H. & Wang, L. Disentangling stereotypes from social reality: astrological stereotypes and discrimination in China. *J. Pers. Soc. Psychol.* **119**, 1359–1379 (2020).
130. Atari, M. et al. Pathogens are linked to human moral systems across time and space. *Curr. Res. Ecol. Soc. Psychol.* **3**, 100060 (2022).
131. Sagi, E. Taming big data: applying the experimental method to naturalistic data sets. *Behav. Res. Methods* **51**, 1619–1635 (2019).
132. Jackson, J. C. et al. Generalized morality culturally evolves as an adaptive heuristic in large social networks. *J. Pers. Soc. Psychol.* **125**, 1207–1238 (2023).
133. Zewail, A., Sepehri, A., Boghrati, R. & Atari, M. Public speakers with nonnative accents garner less engagement. *Psychol. Sci.* **36**, 899–912 (2025).
134. Dale, R., Warlaumont, A. S. & Johnson, K. L. The fundamental importance of method to theory. *Nat. Rev. Psychol.* **2**, 55–66 (2023).
135. Otto, A. R., Fleming, S. M. & Glimcher, P. W. Unexpected but incidental positive outcomes predict real-world gambling. *Psychol. Sci.* **27**, 299–311 (2016).
136. Klein Teeselink, B. & Melios, G. Weather to protest: the effect of Black Lives Matter protests on the 2020 presidential election. *Polit. Behav.* **47**, 1829–1851 (2025).
137. Grzanka, P. R. & Cole, E. R. An argument for bad psychology: disciplinary disruption, public engagement, and social transformation. *Am. Psychol.* **76**, 1334–1345 (2021).
138. Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **3**, 77–101 (2006).
139. Gioia, D. A., Corley, K. G. & Hamilton, A. L. Seeking qualitative rigor in inductive research: notes on the Gioia methodology. *Organ. Res. Methods* **16**, 15–31 (2013).
140. Frith, U. Fast lane to slow science. *Trends Cognit. Sci.* **24**, 1–2 (2020).

Author contributions

G.C.I. researched most of the data for the article and wrote the original draft of the article. F.M.G., L.C. and J.L.T. provided supervision and played a supporting role in researching data for the article. All authors contributed substantially to discussion of the content, as well as reviewing and revising the manuscript before submission.

Competing interests

The authors declare no competing interests. It may be important to note, nonetheless, that L.C. is the Director of the Centre for Gambling Research at UBC, which is supported by funding from the Province of British Columbia and the British Columbia Lottery Corporation (BCLC), a Canadian Crown Corporation. The Province of BC government and the BCLC had no role in the preparation of this article and impose no constraints on publishing.

Additional information

Peer review information *Nature Reviews Psychology* thanks Ashwini Ashokkumar, Olivia Guest, Ori Plonsky and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.