Means to valuable exploration II: How to explore data to modify existing claims and create new ones

Michael Höfler[1,2], Brennan McDonald[1,2], Philipp Kanske[1,2], and Robert Miller [1]

[1]Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

[2]Clinical Psychology and Behavioural Neuroscience, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany

Transparent exploration in science invites novel discoveries by stimulating new or modified claims about hypotheses, models, and theories. In this second article of two consecutive parts, we outline how to explore data patterns that inform such claims. Transparent exploration should be guided by two contrasting goals: comprehensiveness and efficiency. *Comprehensiveness* calls for a thorough search across all variables and possible analyses not to miss anything that might be hidden in the data. *Efficiency* adds that new and modified claims should withstand *severe testing* with new data and give rise to relevant new knowledge. Efficiency aims to reduce false positive claims, which is better achieved if a bunch of results is reduced into a few claims. Means for increasing efficiency are methods for filtering local data patterns (e.g., only interpreting associations that pass statistical tests or using cross-validation) and for smoothing global data patterns (e.g., reducing associations to relations between a few latent variables). We suggest that researchers should condense their results with filtering and smoothing before publication. Coming up with just a few most promising claims saves resources for confirmation trials and keeps scientific communication lean. This should foster the acceptance of transparent exploration. We end with recommendations derived from the considerations in both parts: an exploratory research agenda and suggestions for stakeholders such as journal editors on how to implement more valuable exploration. These include special journal sections or entire journals dedicated to explorative research and a mandatory separate listing of the confirmed and new claims in a paper's abstract.

Keywords: Exploration, Transparency, Smoothing, Filtering, Preregistration, Open Data, Open Analysis, Severe Testing, Replication

Introduction

It has long been recognised that confirmatory and exploratory research are beneficial for each other. Exploratory findings can provide insights for new or improved scientific claims to be tested (Lakatos, 1977; Popper, 1959; Stebbins, 1992), and the failure of a confirmatory trial might suggest exploring for a better claim and a more promising next trial. However, for exploration to inform confirmation well, researchers need to be equipped with an understanding of the aims and means of exploratory analysis in advance.

In the first of two consecutive articles (Höfler et al., 2022), we called for a sharp boundary between confirmation and exploration to separate established from new scientific claims about hypotheses, models and theories. A claim is confirmed if an evidential norm is met, such as p-value (p) $< \alpha$. Strict adherence to an evidential norm ensures severe testing (Mayo, 2018): A confir-

matory test of a claim must be likely to fail if the claim is wrong. Such a risky probe ensures that a claim is supported by meaningful evidence. Unfortunately, adherence is often violated through the use of questionable research practices, by cherry-picking a p $< \alpha$ from numerous different analyses (p-hacking) or a hypothesis that happens to yield such a *p* (HARKing; Hollenbeck and Wright, 2017). Practices like that constitute *intransparent exploration*, misused to produce seeming confirmation of a hypothesis by pretending to meet the norm (Höfler et al., 2022). Behind non-transparency in analysis and generation of hypotheses, non-adherence may be hidden. Therefore, adherence requires control to accept an analysis as confirmatory, for example by pre-registration (Höfler et al., 2022).

In contrast, transparent or "open exploration" (Thompson et al., 2020) enjoys the freedom to extensively analyse data (Manuti & Giancaspro, 2019) and embraces all "researchers' degrees of freedom" (Dirnagl,

2020; Simonsohn et al., 2020) to modify existing or create new claims about the world. However, by trying different analyses, for instance by using multiple statistical tests, the evidential norm may not be adhered to because α accumulates over several tests (Bender & Lange, 2001). In consequence, a confirmatory trial with new data is required to adhere to the norm. This idea extends to concatenated exploration, an iterative process, in which exploration and confirmation repeatedly feed each other, modifying and testing claims, to identify the best possible claims that can be confirmed (Stebbins, 1992). Likewise, empirical science has been described as a process of mapping of knowledge back and forth from a claim via study design to data analysis and modification of the claim, with modification guided, for example, by exploratory results (Bogen & Woodward, 1988; Box, 1980; Lakatos, 1977; Mayo, 2018; Popper, 1959; Suppes, 1969). For transparent exploration to evolve, however, researchers need to be equipped with a conceptual understanding and practical skills of exploratory analysis. This shall foster researchers' self-efficacy and make them more willing to freely conduct and openly report exploration (Stebbins, 2001). However, what exploration actually is has rarely been asked in psychology, with a few exceptions (Behrens, 1997; Dirnagl, 2020). Likewise, exploration, recognisable as such, appears hard to find outside the current data mining/big data movement (Adjerid & Kelley, 2018), qualitative investigations (Kassis & Papps, 2020), planned reviews (Moghaddam, 2004) and theses (Sohmer, 2020).

In this second article we will outline what we believe are important foundations for conceptualising and conducting transparent exploration. We begin with discussing the goals of *comprehensive* and *efficient* exploration. We then describe basic ideas on how to refine existing hypotheses, models and theories and how to create new hypotheses. Based on these foundations, we summarize analytical means to address effectiveness through *filtering* and *smoothing* explorative results. The paper ends with a small research agenda framework and recommendations for stakeholders who have the means to establish more transparent exploratory research.

Goals of exploration
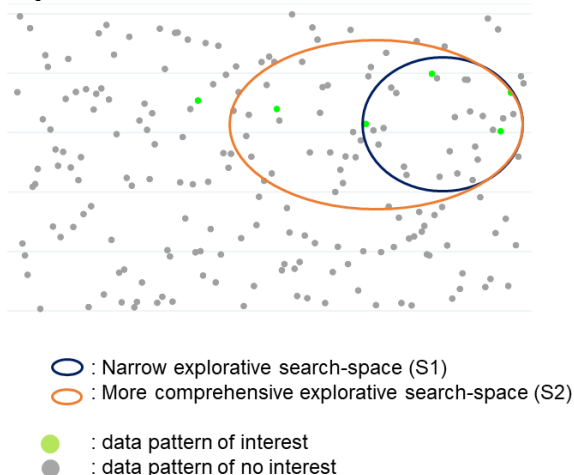
Exploration as a quantitative quest for novelty

As in part I (Höfler et al., 2021), we refer to exploration in the specific sense of "a toolbox of analytical methods to generate and modify hypotheses, models, and theories". Creating and refining such claims about the world allows for scientific novelty and may be achieved by *quantitative analysis*. Note that we do not address qualitative analysis here which may serve the same purpose (I. & R., 1998). We regard quantitative exploration as a quest for *data patterns* that may give rise to novelty. We exemplify data patterns with *associations* between variables, but data patterns may also be higher order relations such as interactions, clusters of individuals or variables that appear similar in a substantive respect, trajectories over time and or other "data regularities" that may point to new insights (Adjerid & Kelley, 2018; Hand, 2007; Nguyen, 2000). A quest for such patterns may be theoretically well informed and thus planned, or may be primarily data-driven, starting with inspection or quantitative analysis of the data resulting in unusual, unexpected or striking patterns. These may be of direct interest or suggest where and how to further explore.

Comprehensive exploration and the explorative search-space

Perhaps the most straightforward idea of exploration is *comprehensiveness*. Comprehensiveness embraces the potential to discover *any and all* patterns in a dataset that would give rise to a hidden truth about nature or challenge prior beliefs (Stebbins, 2001; Swedberg, 2018). Due to feasibility, time, financial and other practical constraints, the resources to explore data will, however, always be limited by the inherent difficulties associated with *collecting new data* or even *analysing given data*. Nevertheless, we suggest that comprehensiveness should initially guide the *planning* of exploration. For example, if one's goal is to identify unknown risk factors for mental health problems, all possible variables, analyses, and observational levels ranging from the biochemical to the level of the society (Williams, 2021) should be taken into consideration in the first place. Theoretical arguments and prior empirical results may then suggest where the most important patterns are hidden. For instance, one may collect a data set with hundreds of potential risk factors from different domains (parental mental health, childhood risk factors, nutrition, stressors ...) and dozens of mental health outcomes (disorders, disability measures ...) and for any factor-outcome combination an association may be found. Alternatively, researchers may decide to focus on exploring a specific domain and a small range of outcomes, for example, diet factors and their relation with affective disorders (Martins et al., 2021). Such considerations ask for boundaries within which to explore. We conceptualize this with the *exploratory search-space*. The exploratory search-space comprises all data patterns (e.g., associations between variables) that are *actually explored* among all patterns that *could* be explored. Choosing an exploratory search-space is akin to

Choosing the explorative search space

○ : Narrow explorative search-space (S1)
○ : More comprehensive explorative search-space (S2)

● : data pattern of interest
● : data pattern of no interest

Figure 1

Schematic illustration of two explorative search-spaces that could be chosen to find data patterns of interest (green dots). A pattern of interest can only be found within the boundaries of a search-space.

placing the lasso of one's practical resources around the area where background knowledge suggests the most novelty. Figure 1 illustrates this with a very simple research quest, where 200 data patterns (potential associations) could be explored, out of which 6 are patterns of interest (e.g., true associations) that could be later identified by explorative analysis.

The figure shows two possible choices of explorative search-spaces, S1 and S2. The narrow S1 contains 4 out of 6 *patterns of interest* and 22 patterns of no interest. S2 is a more comprehensive extension of S1. Exploring S2 requires mores resources. Besides, only 1 more pattern of interest could be identified in a later analysis, but 23 more patterns of no interest could be falsely identified (e.g., by randomly yielding $p < \alpha$).

Efficient exploration

The danger of false-positive results already introduces the second goal of transparent exploration: *efficiency*. Efficient exploration aims to advance science with new insights while not polluting the literature with a multitude of claims. Their subsequent non-confirmation would waste the resources of other researchers in subsequent attempts at confirmation or otherwise fail to advance science. Such findings are the cost of enjoying comprehensiveness in data exploration.

If one turns over every stone, one will find every hidden coin, but also every piece of junk underneath. With "patterns of interest" (the 6 green dots in Figure 1) we suggest that exploration should aim at identifying patterns that are both

1. true and

2. relevant

By "true" we mean that a data pattern is not caused by chance and gives rise to a previously unknown claim, requiring substantive explanation. In the Popperian tradition, a true claim must improve predictions about the world that could turn out to be wrong (Box, 1976; Popper, 1959). Thus (1) aims at finding claims that are likely to *pass severe testing with new data* (Mayo, 2018). Note that a claim derived from a pattern might be close to the pattern; for example, hypothesising an association if a statistically significant association is found in the data. It may, however, require additional substantive input to form a meaningful statement (Rubin & Donkin, 2022). This is strongly the case when a causal hypothesis is derived from the finding of an association (Elhai & Montag, 2020; Glymour et al., 2019).

With respect to (2), exploring for relevant patterns aims to exclude proposals of weak or modest scientific value for the benefit of *stronger* new or modified claims. This serves science per se but also gives rise to more severe testing. A simple example is the claim that an effect is particularly large, rather than just that the effect is greater than zero. Not only is this more scientifically informative, but it can also more easily turn out to be wrong. Generally, with *relevance* we mean any substantive argument that might render a claim scientifically interesting. For instance, causal claims have been argued to be much more relevant than associational claims to inform theories and to assess the potential of interventions (Hernán, 2018; Höfler et al., 2021). Beyond objective dimensions like effect magnitude, practical and clinical significance (Kirk, 1996) or the generalisability of a claim (from a narrow to a more general population), "relevance" is a qualitative term that, we believe, should not be defined in general terms across scientific domains. Perhaps the best general answer to the meaning of relevance is that it must always be renegotiated by the scientific community even within a domain because what appears relevant might itself be subject to change. Note that a wrong claim might nevertheless trigger true insights and thus be in that sense relevant (Nosek et al., 2018; Stebbins, 1992, 2006). For example, claims on ego depletion have not been replicated (Lurquin & Miyake, 2017), but have spawned the idea and finding that willpower is not a limited resource (Job et al., 2010).

Exploring around existing claims

With this understanding of comprehensiveness and efficiency, we are now equipped to derive some basic ideas on how to *actually explore data*. These are not intended to be complete, but rather to sketch out some promising directions that might one day form part of a thorough and mathematically formalized elaboration. We begin with explorative quests around existing knowledge before discussing searches for the entirely new. Thus, we proceed from narrow to wider search-spaces, just as science has been hierarchically classified into single hypotheses, models based on multiple hypotheses, and theories for a full explanation of a phenomenon (Gelman et al., 2019).

Exploring along an existing hypothesis

With specific claims (hypotheses) it is easier to infer what is wrong, while falsifying global claims (models, theories) leaves open which components actually require modification. Additionally, a hypothesis might be wrong but might become true (at least make better predictions) if modified. A hypothesis might also be true but not make a strong proposition. Consider again the *magnitude* of an effect. Commonly researchers hypothesise that an effect is greater than 0 in which case a confirmative result supports any magnitude greater than zero including an effect magnitude arbitrarily close to zero. Thus, an effect could be below any threshold of practical (e.g., clinical, or public health) significance ("nullism", Greenland, 2017). For a stronger proposition, exploration may aim to identify the highest δ, so that the claim "effect $> \delta$" remains true. (Mayo, 2018) gives instances on how to estimate δ based on severe testing calculations.

In general, we believe that "turning all the knobs" (Hofstadter & Dennett, 1981) is a useful metaphor to think about the components of a hypothesis and changing which of them may give rise to a better statement about the world. For example, a hypothesis might state that a particular diet has a positive effect on quality of life. This hypothesis might be modified to say that the effect only occurs in a certain domain of life, or that the diet is only effective if its ingredients are changed. Box 1 describes how trying different analytical methods might lead to a better proposition on an effect or an association.

Box 1: Exploring around a hypothesis with "multiverse analyses" "Specification curves" (Masur & Scharkow, 2020; Simonsohn et al., 2020) and "multiverse analyses" (Del Giudice & Gangestad, 2021; Steegen et al., 2016) try different analytical methods and options and show how a result (p-value, confidence interval) varies across them, how *robust* it is against the assumptions that a particular analysis makes. Then competences on what a method is robust against helps to understand the nature of a relationship under inspection. For instance, there might be clear evidence (p = .001) in ordinary least squares regression for more quality in life on average if a certain diet is followed versus not followed. The evidence might however vanish (p = .450) if "robust linear regression" is used instead, a method that is robust against extreme values and outliers in the residuals (Erceg-Hurn & Mirosevich, 2008; Field & Wilcox, 2017; Huber, 1981; Wilcox, 2012). This may indicate that extreme values dominate the result in ordinary regression if not accounted for. If further data inspection is consistent with that explanation, the initial hypothesis may be refined from a *difference in the mean outcome* to just a *higher probability of extreme values* if the diet is followed. That is, from an overall association to an association only in some individuals. Further exploration, for example with "finite mixture models" (Skrondal & Rabe-Hesketh, 2004), might identify who these are.

Exploring within a theory's or model's degrees of freedom

When modifying a model or theory, "turning all the knobs" calls for questioning all the single propositions from which the model or theory is built. A theory could be broken down into its component parts, changed, if necessary, as described above, and put back together again to form a modified theory. However, it has been criticised that some theories leave knobs unset in the first place, leaving open how they could turn out to be wrong (Bringmann et al., 2022; Scheel, 2021). Underspecification renders them inaccessible to severe testing when tested as a whole, because turning knobs according to the data improves the theory's overall fit to the data (Eronen & Bringmann, 2021; Fiedler, 2017; Gigerenzer, 2010; Lakatos, 1977; Lakens, 2019; Szollosi & Donkin, 2021). Because they are poorly falsifiable, some theories "are not even wrong" (Scheel, 2021). However, with transparency in exploration, filling the gaps becomes an explicit and desirable purpose (Woo et al., 2017). This is rewarded with publications, and a completed model or theory which makes specific predictions and thus becomes subject to severe confirmative testing both as a whole and in its completed

parts. Knobs to be particularly turned are causal claims in theories, which have only been tested as if they were associative (Höfler et al., 2022; Höfler et al., 2021). Another source for *hidden need for modification* is poor measurement with established, but questionable instruments (e.g., Schimmack, 2021).

Exploring to create new claims

Local versus global data patterns

Large-scale studies collect data on many factors and outcomes, such as in the epidemiology of mental disorders (Kessler & Merikangas, 2004), let alone the huge data sets from genetic or imaging studies (Pennycook, 2018; Thompson et al., 2020). With such studies one may find countless associations, and the question arises whether to explore them individually or summarize them in advance (Hand, 2007).

Imagine one assesses 20 nutrition factors in relation with 10 mental health outcomes. Here, *local patterns* are associations between specific nutritional factors and specific outcomes. If indicative of causal effects, they might have *different implications* for science or practice: A theory might suppose that different nutritional factors have very different impacts on various aspects of mental health. Accordingly, interventional effects may depend on which factor is changed to affect which outcome. For example, the absence of alcohol consumption might have a different impact on social well-being than a vegan diet has on personal growth. On the other hand, the 20 factors and 10 outcomes could be manifestations of just a few latent variables, which might explain why a certain set of associations can be found. In this case, one may focus on the *global pattern* of associations, for instance the relation between healthy nutrition and overall mental well-being. Such a focus has been used, for example, to hypothesise about the relationships between psychopathology and neural measures using canonical correlations (Linke et al., 2021). In neuroscience, exploring for global claims has been argued to be more important for insight and prediction (Bzdok & Ioannidis, 2019). Box 2 illustrates how globally focusing on any association versus locally focusing on *specific associations* relates to severe testing when using statistical tests in an explorative manner, and whether one should adjust the α for each test to the number of associations tested.

Box 2: Severe testing of any association versus a particular association if several associations could be found

With 20 factors and 10 outcomes, 200 associations may be tested, each with a level α significance test to separate randomly from non-randomly occurring patterns. Here, α * 200 tests would be expected to yield $p < \alpha$ in the absence of associations (e.g., Colquhoun, 2014). With $\alpha = .05$, this equals 10. If one happens to find at least one $p < \alpha$, the result *"any association found"* has not been severely probed and hence provides only little evidence for anything being truly there, because there were 200 chances for identifying a pattern. Referring to "any association" puts these tests into the global context of all the 200 investigated associations, and with this global perspective, α is inflated (Bender & Lange, 2001). The other possible result, *"no association found"*, would be supported with considerable initial evidence because it could have been 200 times refuted, especially if a sample is large and thus the *beta* errors of the individual tests are small. The evidential norm, however, may be adjusted for the number of tests, α be replaced with α /200 in each test (Bonferroni correction). Not doing so has been criticised to undermine trust in some fields of science through spurious results, for example in genome-wide association studies (Jorgensen et al., 2009; Marigorta et al., 2018). The adjustment turns the matter around: Now the result of "any association" is much more severely probed, but the result of "no association" a great deal less severely than before. If background knowledge suggests that local associations are of interest, each association should be tested with a level α test irrespective of the other associations (Bender & Lange, 2001). A statistically significant association then has been probed with a severity of $1 - \alpha$, and a statistically non-significant association with a severity of 1 - *beta* (Mayo, 2018).

Filtering local data patterns

When searching for the new, it may be desirable to choose a large search-space, but a comprehensive exploration carries the risk of many false positive results. This danger is countered by more rigorous filtering, which results in a smaller number of identified patterns. In Figure 2 we illustrate exploration as a process of firstly choosing an explorative search space as before (Figure 1 with search space S1), then filtering the data patterns and finally creating new claims out of the remaining patterns. After choosing search space S1 with a total of 26 patterns, 4 patterns of interest could be identified. 3 of those are actually identified by filtering, 2 of which being patterns of interest. Then the *efficiency of filtering* within S1 can be described by the proportion

1. Choosing the explorative search space

2. Filtering the data patterns of the explorative search space

○ : Explorative search-space
● : data pattern of interest
● : data pattern of no interest
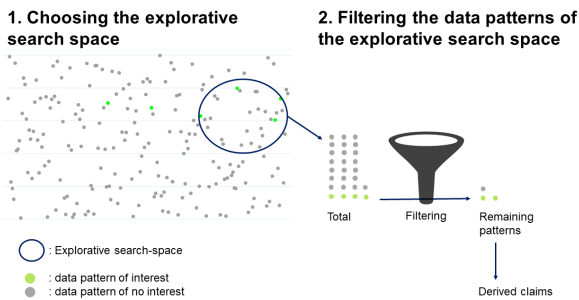
Figure 2

The specification of an exploratory search space in step 1 is efficient in that it covers 4 out of 6 patterns of interest, while the vast majority of patterns of no interest are omitted from the outset (Figure 1). In step 2, the 22 patterns (grey dots) + the 4 interesting patterns (green dots) within the search space are subjected to filtering, after which 1 pattern of no interest and 2 patterns of interest remain. Finally, claims are derived from these.

of identified patterns of interest among all patterns of interest (2 out of 4) and the proportion of patterns of no interest among all identified patterns (1 out of 3). Finally, the 3 identified patterns need to be translated into substantive claims. Most simply and close to the data one may create 3 separate associational hypotheses. Or, 2 associations may appear substantively similar, like: 1. more chronic stress when crash diets are used, and 2. more chronic stress when diet adherence is exceptionally high. This may give rise to more global hypothesizing: "Extreme attention on healthy nutrition is related to more chronic stress".

What are specific methods to filter, here to move from 26 patterns to perhaps 3? So far in this article, we have solely mentioned the dominant method of statistical tests. Statistical tests are useful to eliminate random patterns, but, with their profound origin as an approach to confirmation and without explicit explorative language, they contribute to the blending of confirmation and exploration. This applies as long as the difference is not made very clear by a statement such as "explorative testing was conducted" (Höfler et al., 2022). Alternatives include confidence intervals, descriptive statistics, data mining and machine learning techniques (Adjerid & Kelley, 2018; Alonso et al., 2018; Romero & Ventura, 2020), Bayesian approaches and any other method that may happen to be effective.

Individual versus community-driven filtering

The yet more fundamental question when filtering results is, *who* should do it. With *individual-driven filtering*, as so far assumed, scientists themselves filter their results before coming up with new claims in a publication. The most universal individual filtering method is *internal cross-validation* (De Rooij & Weeda, 2020; Fleming et al., 2021; Xiong et al., 2020). It can, in principle, be combined with any analytical method. Its key idea is splitting a large data set randomly into n subsets, and repeatedly running an exploratory analysis in n-k "training data" subsets while probing its results with the remaining k "test data" subsets (Parvandeh et al., 2020; Xiong et al., 2020). This, however, requires a large total sample size. Most importantly, internal cross-validation endows researchers the freedom to explore *beyond a potentially existing plan* or even *without a plan*, because each pattern found, *no matter with which method and with how many analytical options tried*, must pass the test data. (This works as long as the entire procedure is not repeated with new randomly created subsets until a striking pattern happens to be seemingly confirmed. However, this danger is easy to address by transparency in the seed value of the random process that divides the sample into subsamples.)

By contrast, *community-driven filtering* relies on the scientific public and is usually implemented through per-publication peer reviews. Another instance is external cross-validation, where different data sets are used to generate claims and filter them ("independent replication"; König, 2011). We propose that individually-driven filtering should often precede community-driven filtering, because otherwise rigorously filtered results may receive little attention amidst many published poorly filtered results.

As an important exception to the previous discussions, *each result*, for example on diet – mental health associations, might be potentially informative for other researchers. In such cases, particularly in modestly large search-spaces, no filtering at all seems warranted. All associations may be published on a *public repository* (Pennycook, 2018; Thompson et al., 2020) so that others are enabled to probe the associations predicted from their causal models (Greenland et al., 2004; Ryan et al., 2019).

Smoothing global data patterns

Background knowledge, however, might suggest that one should focus on *global* data patterns beforehand instead of analysing *local* patterns and maybe aggregating these into global claims later (Hand, 2007). With such knowledge one may decide to summarize observed

variables into latent variables before running an analysis, for example by fitting a structural equation model. Or, some entities are known to be more similar than others along a dimension, for example genomic loci along the DNA strand and brain activation or body cells along their two-dimensional spatial distance. Then it is possible to arrange the observations accordingly and to *smooth* the data with statistical methods (Farcomeni & Greco, 2016). Smoothing aims to reduce the variation along the dimension, because otherwise every single point along the dimension is subject to individually occurring random error, potentially hiding the overall pattern of interest or "latent structure". Such smoothing serves to "clean up" the data in the first place (Greenland, 2006).

Consider the example of epigenetic responses to stress exposure across genomic loci. One may explore the variation of the response locally, locus by locus, and thus allow it to freely vary. This preserves *all the patterns* in the data, but many of those will just be noise, the result of random error. The background knowledge that two gene loci are more associated with an outcome the closer they are spatially is ignored (Jaffe et al., 2012). Figure 3 shows a fictive example, in which the outcome **Y**, stress response, varies along the genomic locis' relative spatial location **X** (for illustration one-dimensional and scaled from 0 to 100). The red line displays how **Y** truly varies across location **X** according to the function **Y** = sin(sqrt(**X**)) * 10***X**. We assume that other factors contribute to **Y** through a normally distributed error with expectation = 0 and standard deviation = 500. For smoothing, we use polynomial splines, a technique of non-parametric regression (Takezawa, 2005) that controls the extent of smoothing through the degree of a polynomial (command *twoway lpoly* in Stata 15.2, StataCorp, 2017). Figure 3a shows the random pattern that emerges if no smoothing is done and only local patterns are investigated. Here, the patterns are the spikes that represent outcome values. The height of each spike is a separately estimated parameter. These many estimates (here 50) may poorly carry over to new data, that is, *overfitting* is likely. The spikes might be used to generate a set of individual hypotheses while neglecting the spatial dependency. With luck, such a dependency might emerge with spikes fairly close to the true structure. We suspect, however, that such luck will rarely occur. With moderate smoothing (Figure 3b) we are able to identify a rough course and might hypothesise that the outcome is highest if **X** ranges between 40 and 70. Stronger smoothing (Figure 3c) results in a fairly good fit to the true function and allows further hypothesising a local minimum around **X** = 25. However, if too much smoothing is applied (Figure 3d, linear
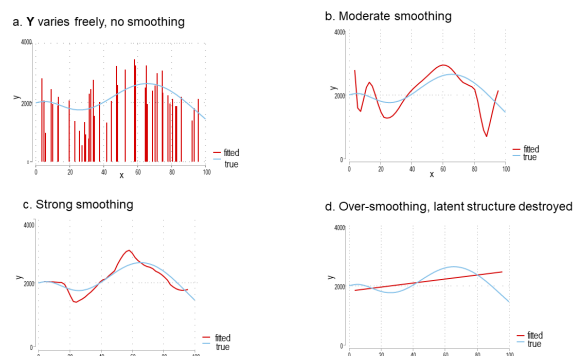
Figure 3

*The figure illustrates fictive data where an outcome **Y** varies across genomic loci (**X**) along a spatial position. The true **Y**-**X** relation (red line) equals **Y** = sin(sqrt(**X**)) *10***X**, where deviations from it arise from random error in a sample of n = 50 (normally distributed with expectation = 0 and standard deviation = 500). Plot (a) shows the results (blue peaks) if no smoothing is done, plots (b) through (d) apply different levels of smoothing, from insufficient smoothing (b) and adequate smoothing (c) to over-smoothing (d).*

approximation)*, underfitting* occurs, the latent structure cannot be described with only two parameters, core features like the peak in the range of 40-70 are overlooked.

Smoothing may reveal a new hypothesis like "stress response has its genetic basis in the range 40-70" or only be an intermediate step (Greenland, 2006), and the smoothed structure (blue curve in the example) be further analysed, e.g., in relation to factors that might influence stress response across location. For example, particularly high peaks in the 40-70 range in individuals with negative childhood events might indicate that genes in this range are activated more strongly in these individuals. Several methods to smooth psychological data are common, albeit not under the idea of smoothing. Table 1 summarises some of them and lists their "smoothing parameters" that regulate the degree of smoothing.

Much elaboration, however, is required for sound guidance on how to apply such methods for exploration: whether they do the right smoothing to the appropriate extent to efficiently stimulate new claims in a specific research domain. As a general advice, the more a field is already understood, the more data may be smoothed.

Table 1

Table 1: Some methods for statistical smoothing, their search spaces and parameters via which they smooth

Method	Search space	Smoothing parameter(s)
Non-parametric regression	Functions that describe an **X-Y** association or the associations of several **X** with **Y**	e.g., the degree of a polynomial (local polynomial smoothing)
Regularisation methods in regression with many predictors (Lasso, elastic net regression, etc.)	Estimates of regression parameters	e.g., the sum of the regression coefficients, besides the intercept (Lasso)
Exploratory factor analysis	Latent dimensions and their loadings on observed items	Number of latent dimensions and choice of rotation method
Cluster analysis, latent mixture models	Possible clusters of individuals that are homogenous within but heterogenous between	Number of clusters
Canonical correlation analysis	Linear combinations of factors and outcomes	Number of latent dimensions behind a set of factors and number of latent dimensions behind a set of outcomes

Planning exploration and transparency on how one has explored

After outlining the goals of comprehensiveness and efficiency and some basic ideas on how to explore data, we are equipped to discuss the possibility of *planning* an explorative quest. We suggest that, if the sample size does not allow for internal cross-validation, a well-underpinned plan may render data exploration more efficient in identifying patterns of interest. The argument is that a planned exploration may be more focused and therefore require less analysis. Identified patterns might in turn be supported by more *initial evidence* (Höfler et al., 2022; Simmons et al., 2011). Note that this is a heuristic argument, because severity depends on what exact analyses are conducted. However, the following strict statement can be made: The severity with which a pattern has been "pre-tested" becomes smaller if *additional* statistical tests are conducted (α becomes greater with each additional test) or any additional filtering has been done.

If there is a plan, it should be *transparent*, that is, made public, to enable researchers to "take credit" of it (Wagenmakers & Dutilh, 2016) when publishing the results that it generates. Also, without a plan, we suggest that transparency beyond the obligatory distinction between exploration and confirmation is crucial for scientific communication. Otherwise, intransparency about how data has been explored could hide some exploratory steps. Readers may then be misled about how promising confirmation attempts are. To give an extreme example, an association might appear present or absent, small or large, positive or negative, just by picking a narrowly defined subsample in which a relation might be claimed (e.g., Vul et al., 2009). If many subsamples have been tried, it may be unlikely that the association will be found again with new data. Box 3 summarizes how some measures inform about what has actually been done and how much initial evidence there is.

Box 3: Transparency measures for what and how much exploration has been conducted"

- **Preregistration**
  Preregistering the pure intention that exploration is to be carried out counteracts later false assertions on confirmation on results that have been actually obtained by exploration (heard from Eric-Jan Wagenmakers in a 2020 talk). If there is a plan on *how* to explore, it should also be preregistered to be later able to show that one had this plan. Changes of a plan might be necessary for various reasons when enjoying the dynamics of digging into data. These can be transparently recorded by an audit trail (version management) system such as "Git" (Chacon & Straub, 2014).

- **Open data**
  Access to data (Isbell, 2021), preferably to the raw data (Arribas-Bel et al., 2021; Nikiforova, 2020; Wilkinson et al., 2016), allows researchers to reproduce found patterns or problems in data that have made changes in a plan necessary. Researchers can also try their own analyses to see if they come up with the same result (Shahin et al., 2020). If such analysis (preferably done by independent re-analysers, e.g., Silberzahn et al., 2018) identifies the same pattern, the initial evidence for the claim is larger, because the alternative analysis might have failed to identify it (e.g., an association that is also found when using a statistical method that is more robust against irregularities in the data such as non-normally distributed residuals; Field and Wilcox, 2017).To ease open access to data, several publishers have recently started to offer purpose-designed, peer-reviewed, and citable journal contribution templates that allow for the publication of data sets (a collection is provided by "Data Journals – Forschungsdaten.org," 2022).

- **Open analysis**
  Open analysis generates transparency in what analyses have been actually done through access to the complete syntax used and the results it has generated (van Dijk et al., 2021). Together with open data it serves reproducing a whole explorative quest. Automatic documentation ensures that no analyses with maybe unfavourable results are concealed. Powerful software packages for this have been developed that store an entire analytical workflow (Peikert & Brandmaier, 2021; Van Lissa et al., 2020; Wratten et al., 2021), as well as notebooks customised for this (Beg et al., 2021).

Further research agenda for exploration: where to explore, what and how to explore

The following proposals summarize the ideas from the first (Höfler et al., 2022) and this second article on exploration. They are likely to give highly context-dependent answers and are therefore intended for separate consideration across the many fields and research quests of psychology. On top of them we invite researchers to *probe* the conceptions of this paper with their own explorative quests. This opens the probably most promising avenue for refinement.

1. *Reconsider established evidential norms for confirmation.* More severe tests may be required to be passed for a claim to move from a new claim to an established claim. In particular, specify what alternative explanations (e.g., bias) must be probed against.

2. *Identify hypotheses, models and theories* that might benefit from exploring around them.

3. *Identify little understood domains.* Specify where key features might be found and with which method of measuring and analysing.

4. What new data should be collected or what existing data should be explored? What are promising search-spaces? Are global or local claims of greater interest?

5. *Identify gaps in theories* that should be filled by exploration to be complete and severely testable. Also identify poorly probed components of theories that might benefit from explorative quests for modifications.

6. *Methodically elaborate on the efficiency of methods of filtering and smoothing.* Consider the applicability of exploratory methods used outside psychology, for example those recommended for exploring the huge amount of medical data in the UK Biobank (2022). Formal elaboration of these concepts should be helpful. Mathematical rigour probes for stringency and may indicate need for change and gaps that have to be filled.

7. Use open science measures for transparency on how exploration was carried out and how much initial evidence may exist for identified patterns.

Recommendations to stakeholders

We end with a list of recommendations for stakeholders including journal editors, peer-reviewers and funding agencies. These three groups have the largest means

for change if they cooperate in addressing the following points. We suggest in general that funding agencies should provide financial incentives for explorative quests, public repositories and methodical elaboration. Editors should offer space and define rules that promote transparent exploration of high quality. Reviewers should control these issues. Open review seems preferable, because it creates transparency in the control process. Specific recommendations are:

1. *Mandatory separation between tested versus new hypotheses* (Gigerenzer, 2018) already listed in the abstract of an article.

2. *Create new journal sections for exploration papers* and reserve space for this (McIntosh, 2017; Thompson et al., 2020). Maybe fund entire exploration journals like the publisher Open Exploration did with its four medical journals (Publishing, 2021).

3. *Use editorials to mention gaps in theories* (Lakens, 2019)(Lakens, 2019) that could be filled by exploration (Woo et al., 2017).

4. *"Place exploratory analyses (regardless of the outcome) on citable public repositories"* (Pennycook, 2018). Funding agencies are requested to create more space and fund according studies to inform other researchers (Thompson et al., 2020) with results suitable to test or feed theories (Greenland et al., 2004).

5. *The common sense that every publication must have an introduction and a discussion part may be questioned.* A pure exploratory publication, for example, on a range of somehow plausible potential risk factors for a disease, does not necessarily require an introduction (it would merely list weak justifications and have little space to describe the theoretical background for analysing each of the many investigated factors). The same applies to the discussion part, a deeper discussion may be better placed in a paper format that discusses the results from several studies and their impact on theory building, interventions and public health (Greenland et al., 2004). Publications on only grossly justified observational data with association results (e.g., short-term planned Covid-19 research) appear most useful if they just describe the methods and report the results (Greenland et al., 2004).

Conclusion

Science has been argued to have made its biggest discoveries through chance (Gaughan, 2010; Roberts, 1989), but maybe chance can be prompted by providing scientists with means to valuable exploration. Psychology seems to have a particularly large potential here. Also, scientific communication could highly benefit from considering exploratory findings not as established knowledge, but as pure suggestions on the rocky path from data to truth that invite one to walk on without knowing where one arrives. Yet teaching some basic insights like how valuable exploration and true confirmation benefit from one another might help, at least in the long run when those who are now taught are ready to conceptualise their own studies. Probably almost every reader has been taught statistics and methods with a nearly exclusive focus on confirmation. Once a new generation of two-trail scientists will emerge, this generation might come up with powerful ways of cooperative exploration that our generation is incapable of imagining because of our confirmatory priming. We wish to conclude with the admittedly emotional remark that the necessity of writing these two articles on the *value of exploration* in science has felt somewhat strange. The self-evidence of this should be reason enough to engage in strict confirmation and transparent exploration and, in turn, to look forward to a science, we believe, thus enriched.

Author Contact

Michael Höfler, Chemnitzer Straße 46, Clinical Psychology and Behavioural Neuroscience, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, 01187 Dresden, Germany. michael.hoefler@tu-dresden.de, +49 351 463 36921

ORCID: https://orcid.org/0000-0001-7646-8265

Author Contributions

Michael Höfler had the lead in developing the conceptions and the writing. Brennan McDonald has contributed epistemic details and was involved in the writing and wording of the entire manuscript. Philipp Kanske commented on and edited the manuscript.

Robert Miller has contributed methodological aspects and reviewed and edited the manuscript.

Open Science Practices

This article earned the Open Materials badge for making the materials openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, *73*(7), 899–917. https://doi.org/10.1037/amp0000190

Alonso, S. G., de la Torre-Díez, I., Hamrioui, S., López-Coronado, M., Calvo Barreno, D., Morón Nozaleda, L., & Franco, M. (2018). Data mining algorithms and techniques in mental health: A planned review. *Journal of Medical Systems*, *42*, 161. https://doi.org/10.1007/s10916-018-1018-2

Arribas-Bel, D., Green, M., Rowe, F., & Singleton, A. (2021). Open data products-a framework for creating valuable analysis ready data. *Journal of Geographical Systems*, *23*, 497–514. https://doi.org/10.1007/s10109-021-00363-5

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiery, N. M., & Fangohr, H. (2021). Using jupyter for reproducible scientific workflows. *Computing in Science Engineering*, *23*(2), 36–46. https://doi.org/10.1109/MCSE.2021.3052101

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131–160. https://doi.org/10.1037/1082-989X.2.2.131

Bender, R., & Lange, S. (2001). Adjusting for multiple testing — when and how? *Journal of Clinical Epidemiology*, *54*(4), 343–349. https://doi.org/10.1016/s0895-4356(00)00314-0

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*, 303–352.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Box, G. E. P. (1980). Sampling and bayes inference in scientific modelling and robustness (with discussion and rejoinder). *Journal of the Royal Statistical Society, Series A*, *143*, 383–430.

Bringmann, L. F., Elmer, T., & Eronen, M. I. (2022). Back to basics: The importance of conceptual clarification in psychological science. *Current Directions in Psychological Science*, *31*(4), 340–346. https://doi.org/10.1177/09637214221096485

Bzdok, D., & Ioannidis, J. P. A. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in Neurosciences*, *42*(4), 251–262. https://doi.org/10.1016/j.tins.2019.02.001

Chacon, S., & Straub, B. (2014). *Pro git*. Apress.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Journal of the Royal Society of Open Science*, *1*, 140216. https://doi.org/10.1098/rsos.140216

Data journals – forschungsdaten.org. (2022).

De Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Del Giudice, M., & Gangestad, S. W. (2021). A traveler's guide to the multiverse: Promises, pitfalls, and a framework for the evaluation of analytic decisions. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920954925

Dirnagl, U. (2020). Preregistration of exploratory research: Learning from the golden age of discovery. *PLOS Biol*, *18*(3), e3000690. https://doi.org/10.1371/journal.pbio.3000690

Elhai, J. D., & Montag, C. (2020). The compatibility of theoretical frameworks with machine learning analyses in psychological research. *Current Opinion in Psychology*, *36*, 83–88. https://doi.org/10.1016/j.copsyc.2020.05.002

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, *63*(7), 591–601. https://doi.org/10.1037/0003-066X.63.7.591

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788.

Farcomeni, A., & Greco, L. (2016). *Robust methods for data reduction*. Chapman; Hall/CRC. https://doi.org/10.1201/b18358

Fiedler, K. (2017). What constitutes strong psychological science? the (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science*, *12*(1), 46–61. https://doi.org/10.1177/1745691616654458

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, *98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.013

Fleming, J. I., Wilson, S. E., Hart, S. A., Therrien, W. J., & Cook, B. G. (2021). Open accessibility in education research: Enhancing the credibility, equity, impact, and efficiency of research. *Educational Psychologist*, *56*(2), 110–121. https://doi.org/10.1080/00461520.2021.1897593

Gaughan, R. (2010). *Accidental genius: The world's greatest by-chance discoveries*. Metro Books.

Gelman, A., Haig, B., Hennig, C., Owen, A., Cousins, R., Young, S., Robert, C., Yanofsky, C., Wagenmakers, E. J., Kenett, R., & Lakeland, D. (2019). *Many perspectives on deborah mayo's "statistical inference as severe testing: How to get beyond the statistics wars"*. Retrieved November 2, 2021, from http://www.stat.columbia.edu/~gelman/research/unpublished/mayo_reviews_2.pdf

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory Psychology*, *20*(6), 733–743. https://doi.org/10.1177/0959354310378184

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

Greenland, S. (2006). Smoothing observational data: A philosophy and implementation for the health sciences. *International Statistical Review*, *74*, 31–46. https://doi.org/10.1111/j.1751-5823.2006.tb00159.x

Greenland, S. (2017). Invited commentary: The need for cognitive science in methodology. *American Journal of Epidemiology*, *186*(6), 639–645. https://doi.org/10.1093/aje/kwx259

Greenland, S., Gago-Dominguez, M., & Castelao, J. E. (2004). The value of risk-factor ("black-box") epidemiology. *Epidemiology*, *15*(5), 529–35. https://doi.org/10.1097/01.ede.0000134867.12896.23

Hand, D. J. (2007). Principles of data mining. *Drug-Safety*, *30*(7), 621–622. https://doi.org/10.2165/00002018-200730070-00010

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

Höfler, M., Scherbaum, S., Kanske, P., McDonald, B., & Miller, R. (2022). Means to valuable exploration i. the blending of confirmation and exploration and how to resolve it. *Meta-Psychology*, *2*(6). https://doi.org/10.15626/MP.2021.2837

Höfler, M., Trautmann, S., & Kanske, P. (2021). Qualitative approximations to causality: Non-randomizable factors in clinical psychology. *Clinical Psychology in Europe*, *3*(2), e3873. https://doi.org/10.32872/cpe.3873

Hofstadter, D. R., & Dennett, D. C. (1981). *The mind's i: Fantasies and reflections on self and soul*. Basic Books.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management*, *43*(1), 5–18. https://doi.org/10.1177/0149206316679487

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc.

I., N., & R., B. C. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Southern Illinois University Press.

Isbell, D. R. (2021). Open science, data analysis, and data sharing. *Open Science Framework Preprint*. https://doi.org/10.31219/osf.io/pdj9y

Jaffe, A. E., Murakami, P., Lee, H., Leek, J. T., Fallin, M. D., Feinberg, A. P., & Irizarry, R. A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International Journal of Epidemiology*, *41*(1), 200–209. https://doi.org/10.1093/ije/dyr238

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion — is it all in your head? implicit theories about willpower affect self-regulation. *Psychological Science*, *21*(11), 1686–1693. https://doi.org/10.1177/0956797610384745

Jorgensen, T. J., Ruczinski, I., Kessing, B., Smith, M. W., Shugart, Y. Y., & Alberg, A. J. (2009). Hypothesis-driven candidate gene association studies: Practical design and analytical considerations. *American Journal of Epidemiology*, *170*(8), 986–993. https://doi.org/10.1093/aje/kwp242

Kassis, A., & Papps, F. A. (2020). Integrating complementary and alternative therapies into professional psychological practice: An exploration of practitioners' perceptions of benefits and barriers. *Complementary therapies in clinical practice*, *41*, 101238. https://doi.org/10.1016/j.ctcp.2020.101238

Kessler, R. C., & Merikangas, K. R. (2004). The national comorbidity survey replication (ncs-r): Background and aims. *International Journal of Methods in Psychiatric Research*, *13*(2), 60–68. https://doi.org/10.1002/mpr.166

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

König, I. R. (2011). Validation in genetic association studies. *Briefings in Bioinformatics*, *12*(3), 253–258. https://doi.org/10.1093/bib/bbq074

Lakatos, I. (1977). *The methodology of scientific research programmes: Philosophical papers volume 1*. Cambridge University Press.

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *PsyArXiv Preprint*. https://doi.org/10.31234/osf.io/jbh4w

Linke, J. O., Abend, R., Kircanski, K., Clayton, M., Stavish, C., et al. (2021). Shared and anxiety-specific pediatric psychopathology dimensions manifest distributed neural correlates. *Biological Psychiatry*, *89*(6), 579–587. https://doi.org/10.1016/j.biopsych.2020.10.018

Lurquin, J. H., & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology*, *8*, 568. https://doi.org/10.3389/fpsyg.2017.00568

Manuti, A., & Giancaspro, M. L. (2019). People make the difference: An explorative study on the relationship between organizational practices, employees' resources, and organizational behavior enhancing the psychology of sustainability and sustainable development. *Sustainability*, *11*, 1499. https://doi.org/10.3390/su11051499

Marigorta, U. M., Rodríguez, J. A., Gibson, G., & Navarro, A. (2018). Replicability and prediction: Lessons and challenges from gwas. *Trends in Genetics: TIG*, *34*(7), 504–517. https://doi.org/10.1016/j.tig.2018.03.005

Martins, L. B., Braga Tibães, J. R., Sanches, M., Jacka, F., Berk, M., & Teixeira, A. L. (2021). Nutrition-based interventions for mood disorders. *Expert Review of Neurotherapeutics*, *21*(3), 303–315. https://doi.org/10.1080/14737175.2021.1881482

Masur, P. K., & Scharkow, M. (2020). Specr: Conducting and visualizing specification curve analyses.

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press. https://doi.org/10.1017/9781107286184

McIntosh, R. D. (2017). Exploratory reports: A new article type for cortex. *Cortex*, *96*, A1–A4. https://doi.org/10.1016/j.cortex.2017.07.014

Moghaddam, F. M. (2004). From 'psychology in literature' to 'psychology is literature': An exploration of boundaries and relationships. *Theory Psychology*, *14*(4), 505–525. https://doi.org/10.1177/0959354304044922

Nguyen, S. H. (2000). Regularity analysis and its applications in data mining. In S. T. L. Polkowski & T. Y. Lin (Eds.), *Rough set methods and applications* (pp. 289–378). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-1840-6_7

Nikiforova, A. (2020). Comparative analysis of national open data portals or whether your portal is ready to bring benefits from open data. *IADIS International Conference on ICT, Society and Human Beings*.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Parvandeh, S., Yeh, H. W., Paulus, M. P., & McKinney, B. A. (2020). Consensus features nested cross-validation. *Bioinformatics*, *36*(10), 3093–3098. https://doi.org/10.1093/bioinformatics/btaa046

Peikert, A., & Brandmaier, A. M. (2021). A reproducible data analysis workflow with r markdown, git, make, and docker. *Quantitative and Computational Methods in Behavioral Sciences*, *1*, e3763. https://doi.org/10.5964/qcmb.3763

Pennycook, G. (2018). You are not your data. *Behavioral and Brain Sciences, 41*. https://doi.org/10.1017/S0140525X1800081X

Popper, K. (1959). *The logic of scientific discovery*. Basic Books.

Publishing, O. E. (2021). *Https://www.explorationpub.com* [Accessed: 2021-01-13]. https://www.explorationpub.com

Roberts, R. M. (1989). *Serendipity: Accidental discoveries in science*. John Wiley & Sons, Inc.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*(3). https://doi.org/10.1002/widm.1355

Rubin, M., & Donkin, C. (2022). Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2022.2113771

Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2019). The challenge of generating causal hypotheses using network models [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/ryg69

Scheel, A. M. (2021). Why most psychological research findings are not even wrong [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/8w2sd

Schimmack, U. (2021). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science, 16*(2), 396–414. https://doi.org/10.1177/1745691619863798

Shahin, M. H., Bhattacharya, S., Silva, D., Kim, S., Burton, J., Podichetty, J., Romero, K., & Conrado, D. J. (2020). Open data revolution in clinical research: Opportunities and challenges. *Clinical and Translational Science, 13*(4), 665–674. https://doi.org/10.1111/cts.12756

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., & et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. https://doi.org/10.1177/2515245917747646

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

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behavior*. https://doi.org/10.1038/s41562-020-0912-z

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC.

Sohmer, O. R. (2020). *An exploration of the value of cooperative inquiry for transpersonal psychology, education, and research: A theoretical and qualitative inquiry* (Doctoral dissertation). California Institute of Integral Studies. https://search.proquest.com/docview/2464456670

Stebbins, R. A. (1992). Concatenated exploration: Notes on a neglected type of longitudinal research. *Quality & Quantity, 26*, 435–442. https://doi.org/10.1007/BF00170454

Stebbins, R. A. (2001). *Exploratory research in the social sciences*. Sage Publications, Inc. https://doi.org/10.4135/9781412984249

Stebbins, R. A. (2006). Concatenated exploration: Aiding theoretic memory by planning well for the future. *Journal of Contemporary Ethnography, 35*(5), 483–494. https://doi.org/10.1177/0891241606286989

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Suppes, P. (1969). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology, and philosophy of science: Proceedings of the 1960 international congress* (pp. 252–261). Stanford University Press.

Swedberg, R. (2018). On the uses of exploratory research and exploratory [Retrieved October 14, 2020].

Szollosi, A., & Donkin, C. (2021). Arrested theory development: The misguided distinction between exploratory and confirmatory research. *Perspectives on Psychological Science, 16*, 717–724. https://doi.org/10.1177/1745691620966796

Takezawa, K. (2005). *Introduction to nonparametric regression*. John Wiley & Sons. https://doi.org/10.1002/0471771457

Thompson, W. H., Wright, J., & Bissett, P. G. (2020). Point of view: Open exploration. *eLife, 9*. https://doi.org/10.7554/eLife.52157

Van Lissa, C. J., Brandmaier, A. M., Brinkman, L., Lamprecht, A.-L., Peikert, A., Struiksma, M. E., & Vreede, B. (2020). Worcs: A workflow for open reproducible code in science. *Data Sci-*

ence, 4(1), 29–49. https://doi.org/10.3233/DS-210031

van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. *Journal of Learning Disabilities, 54*(2), 139–152. https://doi.org/10.1177/0022219420945267

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*(3), 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven selfish reasons for preregistration. *APS Observer, 29*(9). https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Academic Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williams, M. N. (2021). Levels of measurement and statistical analyses. *Meta-Psychology, 5*. https://doi.org/10.15626/MP.2019.1916

Woo, S. E., O'Boyle, E. H., & Spector, P. E. (2017). Best practices in developing, conducting, and evaluating inductive research [editorial]. *Human Resource Management Review, 27*(2), 255–264. https://doi.org/10.1016/j.hrmr.2016.08.004

Wratten, L., Wilm, A., & Göke, J. (2021). Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nature Methods, 18*, 1161–1168. https://doi.org/10.1038/s41592-021-01254-9

Xiong, Z., Chen, Y., Li, Z., & Zhao, Y. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discover using k-fold forward cross-validation. *Computational Materials Science, 171*, 109203. https://doi.org/10.1016/j.commatsci.2019.109203