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

Michael Höfler\textsuperscript{1,2}, Brennan McDonald\textsuperscript{1,2}, Philipp Kanske\textsuperscript{1,2}, and Robert Miller\textsuperscript{1}

\textsuperscript{1}Faculty of Psychology, Technische Universität Dresden, Dresden, Germany
\textsuperscript{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 confirmatory 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,
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

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.

placeing 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 more 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 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 > δ” 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” (Skondal & 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). Under specification 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
Different implications

Specific outcomes. If indicative of causal effects, they are associations between specific nutritional factors and with 10 mental health outcomes. Here, them in advance (Hand, 2007).

Whether to explore them individually or summarize may find countless associations, and the question arises whether to explore them individually or summarize 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, \( \alpha \) 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, \( \alpha \) be replaced with \( \alpha /200 \) 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 \( \alpha \) 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.
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.

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 exposition 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 loci’s 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 $Y = \sin(\sqrt{X}) \times 10^X$, where deviations from it arise from random error in a sample of $n = 50$ (normally distributed with expectation = 0 and standard deviation = 500). 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, plots (b) through (d) apply different levels of smoothing, from insufficient smoothing (b) and adequate smoothing (c) to over-smoothing (d). 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 $Y = \sin(\sqrt{X}) \times 10^X$, 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>
<thead>
<tr>
<th>Method</th>
<th>Search space</th>
<th>Smoothing parameter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-parametric regression</td>
<td>Functions that describe an X-Y association or the associations of several X with Y</td>
<td>e.g., the degree of a polynomial (local polynomial smoothing)</td>
</tr>
<tr>
<td>Regularisation methods in regression with many predictors (Lasso, elastic net regression, etc.)</td>
<td>Estimates of regression parameters</td>
<td>e.g., the sum of the regression coefficients, besides the intercept (Lasso)</td>
</tr>
<tr>
<td>Exploratory factor analysis</td>
<td>Latent dimensions and their loadings on observed items</td>
<td>Number of latent dimensions and choice of rotation method</td>
</tr>
<tr>
<td>Cluster analysis, latent mixture models</td>
<td>Possible clusters of individuals that are homogenous within but heterogeneous between</td>
<td>Number of clusters</td>
</tr>
<tr>
<td>Canonical correlation analysis</td>
<td>Linear combinations of factors and outcomes</td>
<td>Number of latent dimensions behind a set of factors and number of latent dimensions behind a set of outcomes</td>
</tr>
</tbody>
</table>

**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 ($\alpha$ 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.
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 pre-registered 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

**Acknowledgements:** We thank Annekathrin Rätsch for aid with the references.

**Conflict of Interest and Funding**

Robert Miller is an employee of Pfizer Pharma GmbH. The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Philipp Kanske is supported by the German Research Foundation (KA4412/2-1, KA4412/4-1, KA4412/5-1, KA4412/9-1, CRC940/C07).

**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


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*.


