Means to valuable exploration

Michael Höfler

OSIP 19.1.2021,
Klinische Psychologie und Behaviorale Neurowissenschaft, 25.1.2021

Means to valuable exploration: I. The blending of confirmation and exploration and how to resolve it

Höfler M, Scherbaum S, Kanske P, McDonald B, Miller R

Author information

Preprint from **PsyArXiv**, 22 Nov 2021
DOI: 10.31234/osf.io/mtzqj PPR: PPR424983
Preprint

This article is a preprint. It may not have been peer reviewed.

Share this article

Re-submitted (Meta-Psychology)

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

In preparation

Stimulate discussion on open science measures (also relevant to exploration):

- pre-registration
- replication
- open data/open analysis …
- …

Transparent versus clandestine exploration

Transparent exploration:
a valuable toolbox to **modify** existing claims and **generate** new claims (hypotheses, models and theories)

Valuable

p-hacking: clandestine exploration around existing claims, results presented as if they were confirmative

HARKing: (**h**ypothesing **a**fter the **r**esults are **k**nown)
Quest for new claims

QRP questionable research practices

THARKing: transparent HARKing

SHARKing: secretly HARKing to produce seemingly confirmative results

Key ideas

- Degrees of freedom in analysing data are a **curse for confirmation**
  (→ seemingly confirmative results)
- **…** but a **blessing to find novelty**
- Confirmation and exploration are very much **blended**
- Resolve the blending
- … should work better if exploration is embraced as **alternative approach to science**

To succeed, researchers must be equipped with **means to valuable exploration**

Elaboration of conceptions and methods required for
- confidence in one's own exploration
- … and acceptance of other researchers` exploration

and implementation in the publication system

Structure

1. The **blending**
2. **Differentiating** confirmation and exploration
3. **Transparent** exploration serves confirmation and scientific communication
4. **Conceptions** of valuable exploration
5. Exploring around **existing claims** (hypotheses, models, theories)
6. Creating **new claims**
7. Further research agenda for exploration: where to explore, what and how to explore
8. Recommendations to stakeholders to implement more exploration

1. The blending of exploration and confirmation

Blending is not suprising.

Conceptionally:

Analyses on a continuum between

purely confirmatory ↔ purely exploratory

(“where the entire analysis plan has been explicated before the first participant is tested”; Wagenmakers et al., 2012).

(“where the hypothesis is found in the data”)

Teaching: one-sided focus on confirmation → statistical testing misunderstood as ”a universal method for scientific inference” (Gigerenzer & Marewski, 2015)

Evidence for blending

- **Self-reports** on QRP (e.g. excluding data, collecting more data and post-hoc assertions about hypotheses)
- 6 - 11% of researchers across scientific fields even admit to data fabrication and/or falsification (e.g. Gopalakrishna et al., 2021)
- **Content analyses** of papers: almost everything is framed as confirmation and confirmatory (e.g. Banks et al., 2016)
- p-value distributions with peaks just below the usual α = .05 (e.g. Francis, 2012)
- Much more negative results in registered reports (which do not favour positive results; Allen & Mehler, 2019)

Blending **harms scientific communication**:
What is a soundly confirmed claim and what is just a new claim?

What should you ground your research on??

Pressure to produce seemingly confirmative results

- Incentive system: # of publications, impact factors dominate evaluation of scientific performance and career opportunities (e.g. Gonzales & Cunningham, 2015)
- Researchers report such pressure
- … more pressure, more use of QRPs (Gopalakrishna et al., 2021)
- Flawed “ideals of confirmation/telling a good story" (Kerr, 1998)
  → “confirmation bias“, "positive testing strategy" (Klayman & Ha, 1987)
  ↯ Popperian falsification (Popper, 1959)

Pressure partially resists pre-registration

- Records an *a priori* plan of the hypotheses and analyses and thus creates *control* over the plan history (maybe including changes) (Heers, 2020)
- Can be abused by scientists to **sell a result** in case of confirmation success (a negative result would be kept secret; Bian et al., 2020)
- … even possible with registered reports
- Pre-registration is far from being the norm (Beffara-Bret & Beffara-Bret, 2019; Hardwicke et al., 2020)
- Practical issues: Underspecification, non-compliance with pre-registration

- **Unaddressed fear** that research has been in vain in case of non-confirmation?

→ Reduce pressure through means to transparent exploration

2. Differentiating confirmation and exploration

- In sound confirmation a claim must be unaffected by the results
- must use an **evidential norm**
- **…** with a threshold for researchers to be accountable to it (Mayo, 2018)
- … common but disputable: $p < \alpha = .05$

- Pre-registration is the **mode of control on adherence with a norm**
- Accept only pre-registered studies as confirmation!
- Everything else should be considered as exploration

Assessing evidential norms via “severe testing“ (Mayo, 2018)

- E.g. *positive* test result: How likely would the result be *negative* would the claim be wrong?
- If likely, the test has been **severe**.
- → For a severe test, probe a claim rigidly against **alternative explanations** to the claim being true
- (better norms through proper use of confidence intervals)
- Severe testing integrates different statistical schools (Frequentist, Likelihood, Bayesian)
- But Mayo’s elaborations on calculating severity are **themselves frequentist**
- … assuming that alternative explanations besides chance are probed by a study design

- This includes **systematic error** = bias
- No bias means perfect randomized experiment without non-compliance, measurement error, selection effects
- E.g. causal claims (in theories) only tested as if they were associational claims (usually, no causal model is used outside experimental studies; e.g. → shared causes of factor and outcome ignored)
- Probe also against **scepticism** (“sceptical prior”)
- → Bayesian severity may account for bias + scepticism

What is exploration in this respect?

Epistemic price of exploring data:

- Exaggerates amount of evidence (α often inflated through multiple analytic steps)
- The more, the more exploration is done
- → Initial evidence for a claim often insufficient
- (difficult to calculate to what extent; e.g. how much inflation in α)

New data required to achieve an evidential norm.
Extreme case: no trust at all in initial evidence

→ evidence must come entirely from new data
→ replication initiatives (Schimmack, 2018)

Exact **replication** might not probe sufficiently → conceptual / causal replication with new data
(Steiner et al., 2019)

Confirmation is the straight, planned path to scientific insight.

Exploration describes the quests on non-straight paths to find the **unexpected.**

3. Transparent exploration serves confirmation and scientific communication

- With **transparency** questionable research practices are no longer questionable
- E.g. subsample analysis ("… is effective in young female patients")
- (Non-)confirmation is not the end!
- Science is most productive if confirmation and exploration feed one another (Popper, 1959; Lakatos, 1977)
- "**Concatenated exploration**" (Stebbins, 1992): An explorative finding invites a confirmation trial, refinement, confirmation trial, further adjustment …
- E.g. the history of dialectic behaviour therapy (Linehan & Wilks, 2015)
- Researchers who co-operate in such chains may expect citations and a higher probability of **sustainable results**

4. Principles of valuable exploration

Exploration is **valuable** if it advances science by establishing or modifying existing hypotheses, models or theories. Such novelty is triggered by identifying *data patterns*. (Maybe indirectly: Claims might be wrong but anyway trigger true insight.)

Systematic exploration: exploration that follows a plan.

Exploration may also be unsystematic, unplanned; crawl through data in a non-predictable way.

Transparent exploration

- That an analysis was exploratory (rather than conformatory) must always be transparent
- ***How much and what exploration*** has been done to arrive at modified or new claims informs on the initial amount of evidence
- The **exploratory search-space** informs on where and what one has explored
- E.g. associations could have been found between each pair of k factors an l outcomes (k * l possible associations).

Means to transparent exploration

- Pre-registration
- Open data
- Open materials
- Open analysis, automatic documentation tools that store an analytical workflow (e.g. notebooks configured for that)
- … create transparency even in unsystematic exploration

Comprehensive exploration

- At least when planning exploration **consider any option** of where and how to look for the novel
- Enjoy all "researcher degrees of freedom" (Simonsohn et al., 2020)
- In unsystematic exploration: embrace the dynamics that a quest might take on.

Effective exploration

Plan exploration with background knowledge in a way that new claims are expected to be:

- **true*:** separate signals from noise (much noise in huge search-spaces)
- **relevant**:** move science forward through strong statements; e.g. effect > Δ instead of effect > 0

(strong claims give rise to severe testing)

(* meant in a loose sense as a data pattern not caused by chance that reflects or gives rise to something previously unknown, whatever this may be, which requires substantive explanation, whatever the explanation may be. At the very least, efficient exploration makes correct predictions about the world that could turn out to be wrong (Box, 1976))

(** “relevant“ is a qualitative term, its meaning should always be re-negotiated in a particular domain)

5. Exploring around existing claims

- Through modification, a claim should become/maintain *true* and maybe more *relevant*
- "turn all the knobs" (Hofstadter & Dennett, 1981)
- Global claims (models, theories) might turn our to be wrong, but we need to now what components are wrong
- In the Popperian (1959) tradition, focus on **specific claims** that induce specific predictions which could, if wrong, be easily falsified

Exploring around a hypothesis with specification curves or multiverse analysis

- Create variation in results (e.g. p-values) across “theoretically justified, statistically valid and non-redundant specifications”
- Enable “readers to identify consequential specifications decisions” (Simonsohn et al., 2020)

- For example: a group difference is only found with analytical methods that are *not robust against extreme values*
- → modified hypothesis: Belonging to a group does not shift the entire outcome’s distribution (usual assumption) but only increases the *probability of an extreme value*

Other knobs

- Functional form of an association/effect
- An association is there, but “explained away” by certain confounders
- “Outcome shift”: use another outcome
- Thresholds in factor/outcome where an association occurs
- Assess heterogeneity with “finite mixture models”: identify latent classes of individuals across which an association varies
- … relate these to observed variables

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

- Theories often leave knobs unspecified, gaps are filled with clandestine exploration → low severity in testing (Eronen & Bringmann, 2021; Fiedler, 2017; Gigerenzer, 2010; Lakatos, 1977; Lakens, 2019; Szollosi & Donkin, 2021)
- → Some theories "are not even wrong" (Scheel, 2021)
- Hidden need for exploration: Causal relations in theories only probed as associations, e.g. wrong assumption of no bias due to measurement
- Instead use transparent exploration to 1. fill the gaps, 2. test the specified theory (its components)

6. Creating new claims

(focus on hypotheses to build models/theories out of them …)

What are you interested in?

Local data patterns: e.g. association between *specific* factor and *specific* outcome

Global data patterns: e.g. overall associations between a *set* of factors and a *set* of outcomes

One has found an association:

low severity in the global context (it was likely to find **any** association)
high severity in the local context (that a **specific association** might have easily *not* been found)

- E.g. effects* of particular *nutritional factors* on particular *health outcomes* could have **different implications** for science or practice

- Or effects might suggest that factors and outcomes cluster to some *latent variables* that drive the effect.

Closely related to the question whether adjustment for multiple testing is required
(Bender & Lange, 2001)

(* “association“ probably has to be replaced with “effect“ to give this substantive meaning.)

Filtering local data patterns

Individual filtering: researchers themselves filter *before publishing* new claims from data patterns

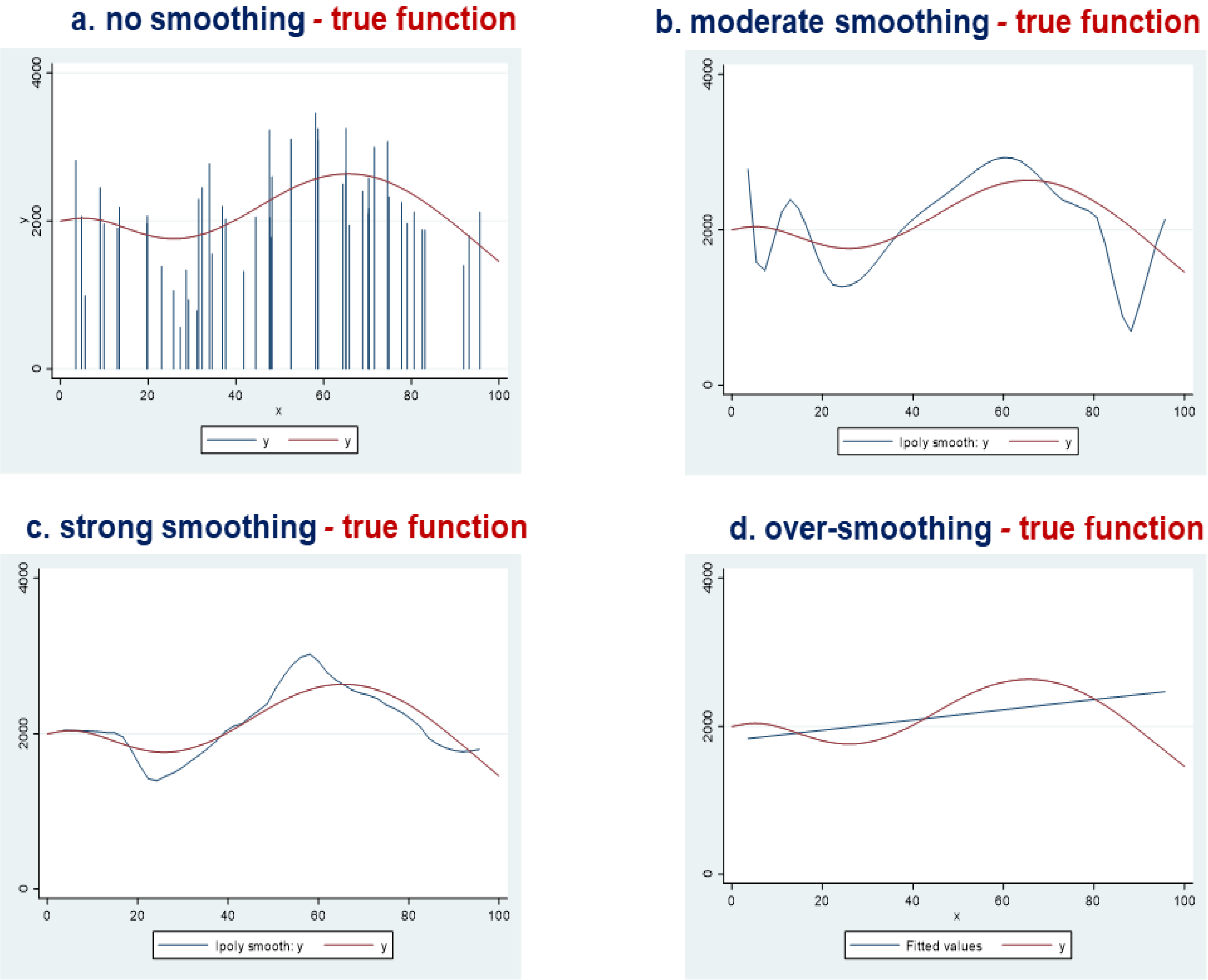
- E.g. (“familywise”) type I error rate when conducting individual statistical tests in genome-wide association studies
- **Internal cross-validation**: Use training data to identify patterns, only maintain those who are confirmed in testing data
- … unsystematic exploration does not necess. reduce initial evidence because this may come from the testing data!

Community-driven filtering: Scientific community decides which patterns deserve attention, most often through peer review
Or all results are of potential interest → **public repositories**

In many explorative quests, individual filtering should precede community-driven filtering.

Smoothing over global data patterns

- Background knowledge: some entities,e.g. genomic loci, are more similar than others along a dimension
- arrange the observations along the dimension (DNA strand)
- *Smooth away random features while preserving latent structures* through avoiding over-smoothing (Greenland, 2006)



*The figure illustrates fictive data where an **epigenetic stress response 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).*

Filtering and smoothing are common in Psychology but sometimes not as such

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

How many hypotheses should be proposed?

- rigid filtering/smoothing strongly protects against random data patterns
- Additional filtering hypotheses with more relevance stimulates more subsequent discussion and research
- a **small number** takes care of the limited resources for confirmatory attempts
- and avoids information overload in the recipients (Buchanan & Kock, 2001)
- And helps acceptance of the hopeful new explorative approach to science

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

1. **Identify hypotheses, models and theories** that might benefit from exploring around them.
2. **Identify little understood *domains*.**
3. **Identify gaps in theories** that should be filled by exploration to be complete and severely testable.
4. **Methodically elaborate** on the efficiency of methods of filtering and smoothing.
5. **Plan and pre-register explorative analysis** as far as possible, be transparent with any changes on the plan; use open data, open materials and open analysis to make the initial amount of evidence assessable.

8. Recommendations to stakeholders
(e.g. journal editors)

1. **Mandatory separation between tested versus new hypotheses** (Gigerenzer, 2018) already listed in the abstract of an article.
2. **Editors should establish control on compliance with the pre-registered**
3. **Create new journal sections/journals** for explorative papers
4. **Use editorials to mention gaps in theories** (Lakens, 2019) that could be filled by exploration (Woo et al., 2017).
5. **Invite distinguished scientists** to provide good examples of exploratory research (Woo et al., 2017).
6. “Place [some] exploratory analyses (regardless of the outcome) on citable public repositories” (Thompson et al., 2020).
7. The common sense that every publication must have an introduction and a discussion part may be questioned.

What is next?

- Invitation to *probe* these conceptions with **your explorative quests**.
- Maybe most promising avenue for their refinement.
- A consecutive paper may provide a practical example: partially planned, pre-registered exploration with open data/open analysis …