An Introduction to Complementary Explanation

Joeri van Hugten [1]

[1] Vrije Universiteit Amsterdam

This paper introduces the practice of complementary explanation; the practice of taking a published result and writing a focused paper that rigorously and systematically describes the implications for a theory that would be rejected by those results. Such spotlighting of a rejected theory counteracts the common alignment between theory and result in published work.

[1]Vrije Universiteit Amsterdam

The reliability of the social sciences is threatened by underreporting. Underreporting refers to reported evidence not reflecting all collected evidence. This is concerning if reported evidence is a systematically disproportionate subset of collected evidence. Currently, this is the case with reported evidence being severely biased toward evidence that supports theories. For example, a large-scale replication of 100 psychology experiments replicated only 36 out of 97 significant results (**Collaboration2015**)(which should be more representative of all collected evidence) are only 33% supported and 42% rejected (**Mazzola2013**; **Hugten2021**)

Underreporting is caused by underreporting practices such as hypothesizing after the results are known (HARKing) and not writing up all conducted tests. This leads to bias because especially results that do not support a paper's theory tend to be the ones not written up, and hypotheses made after the results are known tend to be ones that are in line with those results. Underreporting practices are prevalent. Measuring socially undesirable behavior is difficult, but best efforts suggest that 91% of academics know faculty who engaged in HARKing in the past year, 77% knows faculty who selected data that would support their hypothesis and withheld the rest (**Bedeian2010**; **Rubin2017**)

This paper proposes a practice to challenge theories and counteract underreporting. That practice is based on the falsificationist hypothetico-deductive philosophy (**Witteloostuijn2016**)

Underreporting practices as a neglect of falsification

Besides psychological factors (e.g., confirmation bias) and sociological factors (e.g., not undermining your colleague's theories), beliefs about what is important also underlie underreporting practices. In this section, I speculate about underlying beliefs for three aspects of underreporting practices, as well as a falsificationist principle that speaks to that belief. The posited beliefs are overlapping, and it turns out that falsificationist principles form a coherent opposition to those beliefs.

In the theory and hypothesis sections, why do HARKed hypotheses tend to be in line with the result? My personal intuition is that researchers understand that, at the level of the research program, theories aim to explain phenomena, but that this gets mistakenly transferred to believing that also hypotheses aim to explain results, at the level of the individual study. Given that aim, it follows that hypotheses that are not in line with the result are not useful (**Johns2019**)(as long as they are tightly connected to a theory).

In the results section, why are results against hypothesis the underreported kinds of result? A possible underlying belief is that results that support hypotheses grant more valuable knowledge. In direct contrast, a key principle of falsification is that results against hypotheses give more valuable knowledge. For instance, seminal falsificationist Karl Popper argues that we learn more from results against hypotheses. Broadly speaking, the argument is that a result that supports a hypothesis does not imply that the theory is true, because that is affirming the consequent. Specifically, such an inference would go: 'if theory T is true, then data D should be observed' (i.e., the hypothesis), 'data D is observed' (i.e., the result is in line with the hypothesis), 'Therefore, theory T is true'. This is a logical fallacy because there may be alternative explanations for data D. Therefore, researchers try hard to rule out such alternative explanations by using random assignment, control variables, or more advanced statistical techniques. By contrast, a rejected hypothesis does imply that at least one premise in the theory or operationalization is false, because it is denying the consequent

Table 1 Possible beliefs underlying underreporting practices.

Underreporting practice	Possible underlying belief	Related falsificationist principle
HARKed hypothesis tend to be in line with the result	Theories aim to explain the world, therefore hypotheses aim to explain the result	Hypotheses aim to challenge theory
Unreported results tend to be against the theory	Supported hypotheses give more valuable knowledge	Rejected hypotheses give more valuable knowledge
Discussion section focus on explaining the result	Explaining the result is the goal of a paper	Challenging the theory is the goal of a paper

Table 2 Complementary explanation in relation to existing practices.

Focus on before the result is known	
Focus on the result	
Focus on the rejected hypothesis	Counterargument, co

which is a valid form of argument (even if alternative explanations were not excluded). Less extremely, (**Davis1971**)

In the discussion section, if a rejected hypothesis is reported, why is the expectation that authors explain the result? I speculate that the underlying belief may be that explaining data is a more important goal than improving theory. By contrast, falsificationist principles hold improving theory as the main goal. Therefore, those principles suggest that discussion sections build on the result to contribute contingencies that make the theory less simple or generalizable, and as a result, more accurate (**Lakatos1970**; **Cross1982**). Contributing contingencies can also happen in the process of explaining a result. However, the distinction is especially clear when discussion sections bring in a completely different theory that does fit the result. The distinction also becomes clearer if one imagines a more extreme alternate world in which discussion sections purposefully attempt to bring in additional theories that are opposite to the result. By contrast, current practice is that no further discussion is needed once the result is explained.

Overall, the argument is not that following the principles of falsification leads to more ethical research; it probably only affects the type of results that are underreported, not the extent of underreporting. That is, if researchers believed that rejected hypotheses lead to more valuable knowledge, then underreporting might start tending toward underreporting results that support hypotheses. Currently, the tendency is to underreport rejected hypotheses, so a practice based on principles of falsification can help bring balance.

A proposed counteracting practice: complementary explanation

Because of the opposition to falsificationist principles in the aspects of underreporting practices, I propose that a practice that thoroughly applies those falsificationist principles can counteract underreporting practices. Specifically, I propose complementary explanation (CE).

The term 'complementary explanation' is a variation on the term 'alternative explanation'. An alternative explanation is an explanation for a result and an alternative to the hypothesis development (assuming that the result was in line with that hypothesis). Alternative explanations are the main threat that Popper aimed to avoid. By contrast, a complementary explanation (CE – countable) is an explanation for the opposite of a result, so it is a logical complement to the hypothesis development (assuming that the result was in line with that hypothesis). For example, if a quantitative study finds a positive coefficient, a CE for that result is a set of arguments that imply a negative coefficient. Similarly, for a qualitative study's causal story between high X and high Y, a theory's implication of a negative relation is a CE. If a study's result is inconsistent with its hypothesis, then the original hypothesis development is a CE. Even if a study does not have a hypothesis for a particular relation, an explanation of the opposite of its result is a CE. One result can have multiple CEs.

To appreciate CE's unique focus, table 2 positions CE in the context of a comprehensive list of similar existing practices. CE is similar to counterarguments, competing hypotheses, meaningful baselines, or theory-driven null hypotheses (**Schwab2012**)

CE is like HARKing and spinoffs like Tharking (i.e., transparently hypothesizing after the results are known (**Hollenbeck2017**; **Rubin2017**) and abduction (**Locke2008**; **Schwab2017**) in that all those practices happen after a result is found. However, the difference is that those practices aim to explain a result (although RHarking is, in principal, also open to rejected hypotheses (**Rubin2017**)). For example, abduction would never involve explaining the opposite of the result. In other words, hypotheses made after the results are known tend to be ones that are in line with those results. But they need not be that way. CE is like transparently making a hypothesis after the result is known, that is opposite to that result. That shift in focus counteracts the threat of HARKing to research reliability. Finally, CE is like SHarking (suppressing hypotheses after the results are known); the most threatening form of HARKing (**Rubin2017**), except that SHarking focuses on suppressing rejected hypotheses while CE adds exactly such hypotheses.

Table 3   *CE Steps.*

Steps	Notes
0. Find a result	While reading, one might stumble upon a published finding that...
1. Develop a CE for that result	What collection of premises suggest the opposite of the result? By design, the CE is not in line with the result. So at least one of...
2. Identify a premise in that CE to challenge	What would... a CE paper except that CE complicates...
3. Suggest a complication for that challenged premise	Does the complication increase accuracy? Is the complication pl...
4. Evaluate that complication's effect on accuracy.	The most simple and generalizable complication that can accura...
5. Iterate over steps 3-4. When out of ideas, compare complications.	The less paper-specific the challenged premise, the greater the t...
6. Iterate over steps 2-3-4-5. When out of ideas, compare challenges.	Concisely and concretely describe the new insight. E.g., 'Premise...
7. Specify the contribution	

Note: While X is typical. For example, see **Cook1979** *discussing why insignificant results regarding the cognitive bias 'sleeper effect' are due to operationalizations not appreciating theoretical nuances. Cook et al.'s (1979) paper would like a CE paper except that CE complicates the discussion from 'in this paper we remind people of some important nuances in the theory, which empirical studies have failed to appreciate, which led to insignificant results' to 'insignificant results have forced us to appreciate the theoretical nuances, and in this paper, we make explicit the nuances we now believe to be important'.*

CE Steps

The steps to interpret supportive results seem clear: e.g., 1) p<0.05, 2) hypothesis supported, and 3) more confidence in the theory (but see **Wasserstein2019** for how it is not that simple). By contrast, the application of falsification is impeded by a lack of such clear steps. CE is a way to codify falsificationist interpretation steps. Table 3 summarizes these steps.

A crucial step in falsification is that a rejected hypothesis implies that at least one premise in its explanation is false. But, it is undetermined exactly which one is false (**Hines1988**; **Lakatos1970**; **Søberg2005**; **Weick1999**)

The fact that such trade-offs are possible to 'save a theory from falsification' has been used to argue against falsification (**Søberg2005**)(or research programs) with many such trade-offs are degenerate (**Lakatos1970**; **Cross1982**)(a research program in macroeconomics) while explicitly reflecting on the Lakatosian ideas at the basis of that judgment.

Step 0 and step 1 contribute by identifying a lack of accuracy. Step 0 may seem difficult, but the same creativity that is displayed in thinking of alternative explanations should also allow us to reinterpret results from theories that oppose that result. Regarding step 1, developing a CE does not require fully fleshed-out theories. Instead, CEs consist of the most straightforward, and commonly held, associations of concepts (**Davis1971**; **Spector2011**)

Steps 2 to 6 contribute by identifying ways to restore accuracy by trading-off simplicity and/or generalizability. Thus, one of the accepted propositions is negated and replaced by a proposition that is more complex and more 'interesting' (**Davis1971**)(and the result was found for a solid planet). CE encourages including that challenge, even if another challenge ends up being more plausible.

Step 7 makes summarizes the complication; making explicit the degeneration that is forced upon the theory by the result. It is possible that a result is inconsistent with a CE because of bad measures, auxiliary premises, or research designs. The CE author can decide whether a CE with a step 7 that reads 'Measure X does not capture concept A (in some context)' contributes enough to be worth the effort. If a challenged premise is paper-specific, the contribution may be small. On the other hand, the contribution may be large enough if measure X is typical. For example, see **Cook1979** discussing why insignificant results regarding the cognitive bias 'sleeper effect' are due to operationalizations not appreciating theoretical nuances. Cook et al.'s (1979) paper would like a CE paper except that CE complicates the discussion from 'in this paper we remind people of some important nuances in the theory, which empirical studies have failed to appreciate, which led to insignificant results' to 'insignificant results have forced us to appreciate the theoretical nuances, and in this paper, we make explicit the nuances we now believe to be important'.

Full circle

Given those details, we can see how CE counteracts underreporting by improving meta-analyses. Meta-analyses use concept labels as inclusion criteria. For example, **Heugens2009**'s meta-analysis on 'mimetic pressure's effect on isomorphism' uses a variety of concept labels to search for literature (e.g., 'isomorphism', 'institutional theory'). Underreporting practices cause studies to be described in terms of concepts that are supported by the result. Therefore, meta-analyses disproportionately include studies that support the theory (**Murphy2019**). (Note also that while Tharking (**Hollenbeck2017**) and abduction (**Locke2008**; **Schwab2017**) do not mislead like HARKing, they still lead studies to be described in terms of concepts that are supported by the results).

Enter CE. For example, a study finds a positive effect of 'competition' (measured as the number of firms in the same industry) on 'differentiation'. A CE for that result is that a greater number of firms in the same industry can be interpreted as mimetic pressure (**Haveman1993**)(systematically) fail to include results like these opposite to the theory. By contrast, after the CE is published, the result is described using theories that are not supported by it, so meta-analyses would include it. HARKing and not writing up tests could continue at the usual rate, but with CE, the proportion of rejected hypotheses among reported evidence would be greater (and closer to the true proportion).

Conclusion

Proposals to combat underreporting focus on preventing underreporting practices; e.g., de-emphasize p-values (**Bettis2012**; **Bettis2012**; **Wasserstein2019**; **Schwab2011**)

Moreover, CE is a useful practice, even if underreporting did not exist. First, CE also increases research reliability more directly. When CE is done by others than those who found the result, research reliability is increased simply by having an extra person thinking through the meaning of the data from

a fresh perspective. Second, we put strain on others when collecting data. This comes with a responsibility to make the most of our data. CE helps fulfill that responsibility by reusing published results, in contrast to demands for efficient rather than comprehensive presentation, and novel findings.

In sum, I hope people use CE to learn more from the same findings and especially learn about, and from, those things that we currently miss due to underreporting.

Acknowledgements

References