Why Are Most Published Research Findings Under-Theorized?

Or: Are We in an Interpretation Crisis?

Benjamin Krämer

Abstract
While the so-called replication crisis is increasingly discussed and addressed through reformed research practices and institutional structures, this contribution diagnoses a theory or interpretation crisis and argues that the current emphasis on transparency, reproducibility, and reliability should be complemented by stronger efforts in terms of theory and validity. The article identifies different types of unsystematic (e.g., ad hoc, asymmetric, or trivial) theory building. Furthermore, objects of investigation, measurements, and findings are not interpreted carefully in the light of sufficiently elaborate and well-justified theoretical concepts or frameworks. Different consequences of such shortcomings are discussed—whether unfruitful or implausible hypotheses are tested or the implications of findings are remain poorly understood or are not critically reflected. Readers are invited to engage not only with methodological literature and previous research, but also with theoretical works and particularly with literature on methods or strategies of theorizing, and to practice theory building based on a clearer understanding of the multiple meanings and functions of theory.

John P. A. Ioannidis is probably most known for his article provocatively entitled "Why Most Published Research Findings Are False" (Ioannidis, 2005). He not only provides a strictly statistical argument that, all other things being equal, only a certain proportion of significant results can actually be true if a given Type I and II error rates are accepted (i.e., if the limit for p values and the power of a test are set at a certain level). He also mentions a number of corollaries that, somewhat contrary to his claims, cannot directly be deduced from his main argument. I will criticize one of them later.

I consider John Ioannidis a bit of a tragic figure of science and his story as a cautionary tale: He became famous by fighting for more trustworthy

science through sufficient sample sizes, systematic reviews and meta-analyses, replication, the avoidance of data dredging, and other means that increase the probability that a positive finding is indeed true. But he later became a hero of those who wished to downplay the seriousness of Covid-19, thus effectively undermining trust in other scientists and public policy (that are, of course, never above criticism, but his conclusions were considered problematic based on the best epidemiological evidence, methodological principles, and normative arguments). This was possible because abstract methodological rigor cannot substitute substantial ideas about the object of inquiry, the reflection of ideological biases, or a clear conception of potential distortions of a study that cannot easily be corrected statistically or by means of replication in other contexts. Such biases can only be discovered by theorizing the object of inquiry and arguing about the validity of the methodology.

Fittingly, the article that Ioannidis co-authored and that demonstrates this lack of proper argumentation involved social media (Bendavid, Mulaney, Sood, Shah et al., 2021). The paper arrived at an unusually low infection fatality rate of Covid-19 in comparison to other studies, was widely publicized as an argument against strict containment measures, and was criticized for all kinds of reasons, ranging from details of the statistical analysis to potential undisclosed conflicts of interest (Lee, 2020). Most importantly in our context, the authors devote a lot of space and the most complex meta-analytical procedures to the calculation of the sensitivity and specificity of the antibody test. However, they only shortly comment on the bias due to participant self-selection in a revised version of their manuscript and initially only weighted their data along selected sociodemographic categories. Only in the later versions, they present some back-of-the-envelope calculations based on reported symptoms that are meant to address self-selection. However, a thorough discussion of who will probably respond to a Facebook ad inviting users to be tested after a drive to a test center would have shown that the bias of estimates based on such a curious, mobile, flexible etc. population is essentially incalculable. One does not have to call such a discussion a "theory" of recruitment via Facebook or of voluntary antibody testing (but why not? Let's not put the bar of what constitutes theory too high… However, a lot of research on social media and current media change *could* indeed profit from more elaborate theory in order to better assess how new developments affect all kinds of fields and activities, such as health or social research). But the total lack of such a reflection in the initial version of the paper shows that substantial ideas on an object of investigation or on participant behavior

cannot be replaced by the most sophisticated methods from the toolkit Ioannidis rightly recommends.

Over the years, I have reviewed and otherwise read many manuscripts in different areas of quantitative communication research with elaborate designs and sophisticated data analysis but with a number of recurring problems: Sometimes, the setting of the study, the wording of certain items, the style and content of a stimulus etc. did not really correspond to any real-world setting (even with some concessions that would be necessary for methodological reasons). In other cases, it was not clear whether the researchers had successfully manipulated or measured certain constructs given the ambiguity of certain concepts or of certain wordings in questionnaires. Or it was unclear whether they had successfully demonstrated a causal link because important confounding factors or biases had been overlooked.

These different cases constitute problems of validity—either external validity or the validity of causal inference and measurements. In both cases, this is not a problem of methodology as such but of the specific conceptualizations and theories of the object of investigation. Whether a study can claim to model or mimic a real-world situation can only be determined if we have a sufficiently clear idea about this type of situation. Whether a study can claim to have measured or manipulated a given construct or to have demonstrated a causal link without being misled by spurious correlations or biases can only be determined by a clear idea about what constructs and words or phrases mean and how phenomena are related.

I have also often noticed further problems beyond validity proper. Sometimes, the postulated causal order seemed arbitrary, or the explanation of hypotheses careless and unsystematic. I felt that instead of A causing B, it could easily be the other way round (or there could be no real causal relationship), or that instead of the explanation that more A causes more B, it may as well cause less, or instead of A being the main cause of B, it may be C, but A was picked for no particular reason.

Finally, I have also often found the interpretations of the results, the discussion of potential implications, and the reflection on limitations rather trivial or only loosely connected to the specific study—either extremely generic or overly narrow, without regard to the overall social context or scholarly discourse.

Of course, readers may brush this broad criticism of the field aside and assume that I only want to look down on some supposedly narrow-minded colleagues or that I am disappointed because my idiosyncratic pet theories are not reflected in current research. On the contrary, I would argue

that there is number of specific reasons why theory building (including epistemological and methodological reflection beyond the technicalities of specific methods) is not always satisfactorily systematic and elaborate in mainstream communication research (see also Hagen, Frey & Koch, 2015, for similar observations). And I would further argue that this has a number of specific implications.

In sum, the argument will be that we are not only in a replication but an interpretation crisis, a crisis of theory building, and that the (legitimate) focus on transparency, reproducibility, and reliability (e.g., with open science, replications, meta-analyses etc.; Dienlin et al., 2021; Rains, Levine & Weber, 2018) should be complemented by stronger efforts in terms of theory and validity. To put it provocatively: If we do not know what our findings mean, there is no use reproducing them (or we do not even know what would count as an actual replication—see below). In psychology, it has even been argued that a theory crisis is one of the factors explaining the replication crisis because hypotheses that are not well-justified theoretically are more likely to result in false-positive findings and because the less explicitly and precisely a theory is spelt out (or the less it is clear what would actually count as valid measurement of the relevant constructs), the more difficult it is not decide what would count as an (un-)successful replication or what would even count as a test of a theory versus discovery-oriented research around a loose theoretical framework (see Eronen & Bringmann, 2021; Gigerenzer, 2010; Oberauer & Lewandowsky, 2019).

And the development or application of increasingly elaborate methods is not always matched by highly sophisticated or even sufficiently elaborate theories. Again provocatively: If we do not know what we are looking for or what our results mean, there is no need for complex methods (actually, no need for data collection and analysis of any kind). Ultimately, confronted with several planetary crises and threats to democracy, this is not the time to accumulate data without a clear idea what is or could be going on.

What Kind of Theory?

It should have become clear that this contribution focuses on quantitative research and its logic (usually variable-based and often focused on causal relationships). This is not to say that the state of theory building around qualitative research is beyond criticism. Even though there is also rather atheoretical qualitative research, some approaches to qualitative research explicitly focus on theory building, such as grounded theory. Due to the potentially complex interplay between theory and empirical studies, qua-

litative research has been discussed as "Theoretische Empirie" (Kalthoff, Hirschauer & Lindemann, 2008), 'theoretical *Empirie*,' the latter being a mass noun with no exact equivalent in the English language referring to everything empirical or the totality of empirical activities or findings in a given context. However, the role of theory in qualitative research is simply beyond the scope of the present article. It is an invitation to avoid bad theory and to do good theory, addressed at quantitative researchers in communication science. Although the focus on theory building as the crafting of sets of interrelated testable propositions can be criticized (see below), there is a considerable number of publications on how to establish such and related theories (e.g., Abbott, 2004; Bell, 2009; Elster, 2007; Jaccard & Jacoby, 2020; Runciman, 1983; Shoemaker, Tankard & Larsorsa, 2004; Sohlberg & Leiulfsrud, 2017; Swedberg, 2014b; Stinchcombe, 1986) and on other ways of stimulating theoretical thinking in the social sciences (the "sociological imagination," Mills, 1959, the "tricks" of social-scientific thinking, Becker, 1998, or the conditions for "intuitive theorizing," Knorr-Cetina, 2014). Therefore, the conditions for building theories of this type are relatively good and the need for it should have already become clear and will be substantiated in the following.

However, testable hypotheses cannot easily be separated from other types of theoretical statements.[1] Those include interpretive, conceptual, ontological, epistemological, and methodological statements—sometimes combined into whole "worldviews" and sometimes as solutions to specific problems—as well as normative theories of society and exegeses of classical works (Abend, 2008; Büttner, 2021; Merton, 1945). In other fields, one may debate whether certain exegetical exercises and esoteric analyses of minutiae of certain meta-theories are still fruitful for the understanding of social phenomena—or for the discussion of current challenges of representation and perspectivity (Krause, 2016). The challenge in communication research is probably different: to develop *better* explanations (and better concepts), but also to develop a greater sensibility for other types of theories that are closely connected with explanations (such as conceptualizations and epistemological frameworks) and that fulfill important other func-

1 Certain authors such as Abend (2008) emphasize that they are distinguishing *meanings* of the word "theory," not *types* of theories, because it is difficult or impossible to provide a single definition of the unitary concept of "theory" of which there could be different types. However, what is identified as the different meanings of "theory" in such semantic analyses, can at least be combined into a single "theory" or theoretical work and there are often necessary connections between the different types of statements.

tions (such as the justification of normative judgments or the reflections of judgments that are already unintentionally implied in seemingly value-free explanations).

Theoretical statements form continua or systems where one type of statement cannot easily be separated from another or where no single line can be drawn between "theory" and "non-theory" on the spectrum ranging from general presuppositions and specific observations (Alexander, 1982, p. 2f.). Even if one part is not spelled out, one can still ask what the parts that are not explicated may be (i.e., the more abstract presuppositions of single hypothesis or the empirical fruitfulness of a general theory). Thus, it makes sense to systematically reflect on the whole range of implications of one's explanations or hypotheses instead of developing them *ad hoc* and without any broader theoretical context.

For example, many arguments around newer online media or practices do not only involve individual hypotheses but are embedded into a set of postulates of historical trends and normative judgements (both often used to establish relevance, for example: Social media were once hoped to democratize *X*, but they did not), a set of ideas about the structure of society and social or psychological ontology (for example, a worldview in which there can be political *systems* and media *organizations* that can somehow respond to *technological* change, *actors* with *attitudes* that are *influenced* by new types of messages such as comments, etc.), epistemological assumption (e.g., about the validity of self-reports or the feasibility of automated analysis of meaning), a set of individual interpretations of entities and observations in terms of concepts (e.g., that Facebook and Google are both important new *intermediaries* in today's media environment), etc. If all of these types of conceptions are not carefully reflected, contradictions and confusion may arise (e.g., from category errors such as equating organizations with the sum of their members), problematic myths about society and history risk to be perpetuated (e.g., that there once must have been an era when everyone was ready to compromise, trusted the established media, and always sticked to facts in political discourse), and research designs may fail due to faulty assumptions.

The Problems with Bad Theory

Atheoretical research is easy to criticize, because "letting data speak for itself" means to rely on theory-like preconceptions and ad-hoc interpretations—or not to understand certain findings at all. One can sometimes see people put off a thorough discussion of the possible relationship between

variables until the statistical results are in, only to realize that they still do not know why the relationship they found should exist and what to make of it.

In some cases, it can be acceptable to keep it simple, to rely on a number of everyday categories, and investigate how certain things are related. One simply needs to make sure to avoid the most blatant misconceptions and biases. In this case, our everyday understanding is already a sufficiently good "theory." However, I have observed certain styles of explicit and *ex ante* theorizing that does break with everyday conceptions in favor of more scientific terminology and concepts but that can nevertheless be dangerously biased or simply much less fruitful for our understanding of the social world than it could be. Thus, I will argue that theory building is not only a way of developing new ideas for research and something that we can dispense with if we can still come up with new studies. Nor is it something that we only need if we do not immediately understand our object of study or that only needs to be "just good enough" to make everything somewhat plausible. Instead, good theories (or at least thoroughly checked everyday conceptions) are a general prerequisite for the validity and usefulness of findings. However, a number of problematic strategies of theorizing in quantitative communication research (discussed below) sometimes prevent scholars from developing good explanations and understanding their phenomena well.

One might argue that bad theory work is harmless because it only concerns the "context of discovery" which is irrelevant to the actual research process, the testing of hypotheses, in which bad hypotheses are thrown out anyway. In a somewhat more nuanced manner, Popper (1935) held that the discovery of hypotheses is an irrational process that may be analyzed psychologically but is not completely open to rational reconstruction or justification, and his systematization did not include the process of theoretical justification of hypotheses (for an even more nuanced discussion of the different views on the distinction between the "context of discovery" and "justification," see Hoyningen-Huene, 1987). He only saw room for four types of tests: 1. for consistency, 2. for tautology (or falsifiability), 3. whether a theory postulates something new in comparison to older ones, and 4. testing through empirical application (Popper, 1935, p. 6).[2]

2 If this view were taken absolutely seriously by researchers claiming to be critical rationalists in the tradition of Popper, "theoretical" sections of manuscript in the social sciences would look quite differently or may not even exist in the current form. They would merely mention the hypotheses, maybe dispel any doubts that they are inconsistent internally or among each other, or argue that their

However, there are still theories that pass the first three tests that I would consider bad theories, for example because they are implausible, even if they are not strictly incompatible with well-established findings.

The view that bad theory development is harmless because of the fourth test would be naive for two reasons (that are of interest here, being only two of many reasons why the strict separation between discovery and testing is problematic). First, testing bad theories is a waste of resources. Although according to a Popperian logic, improbable hypotheses are very informative if they are not immediately and convincingly refuted, testing theories that are most probably doomed from the outset is to set problematic priorities. Certainly, science is about curiosity, even personal and collective fulfillment, and about the pursuit of knowledge without any foreseeable practical applications. And as problematic as populist criticisms of scientists "wasting taxpayers' money" are—science is practiced with limited resources and there is a certain obligation to focus on fruitful avenues (not necessarily the ones that promise the highest return on investment in the strict sense, but those that have a sufficient chance to produce actual insights). Furthermore, a seeming confirmation of a hypothesis that is *a priori* very likely to be false is more likely to be a false positive, as Ioannidis (2005) demonstrated. Thus, time for theorizing and developing and selecting the most plausible hypotheses is time well spent.

But more importantly, second, bad theories actually lead to bad interpretations of findings, with actual problematic consequences. Indeed, more or less naive falsificationism does not have a problem with bad theory because it is assumed to be in risk of being refuted by empirical findings. However, hypotheses cannot be considered in isolation. They are interwoven in a network of meanings, assumptions, logical rules, etc. According to more elaborate theories of "holist underdetermination" (Stanford, 2021)[3], such a network cannot be discarded at once by empirical findings. If findings seem to contradict one statement in such a network

---

testing will yield any important insight in comparison to previous research. Strictly speaking, there would be no need to justify or "deduce" the hypotheses, for example by drawing analogies with findings on similar phenomena, as is often done. The papers would then focus on the demonstration that they hypotheses have been tested rigorously, and on the results.

3 "Underdetermination" refers to the idea that theory is underdetermined by data, i.e., that available evidence never completely determines which theory or modifications thereof we should commit ourselves to, never allows us to pick exactly one (new) theory that would be the only one to match the data. "Holist" underdetermination refers to the argument that statements in theories are related and that therefore, evidence alone cannot determine what modifications in a theory should

(and even if there is no such contradiction), there are always multiple alternative arrangements that can in principle be accepted as consistent with the data (and a contradiction can only be asserted based on other parts of the network that convey meaning to the statement and that connect it to the data). If a mismatch is identified, we can always question the method of data collection, the deduction of specific testable prediction from general relationships, the meaning of statements implied in all steps of the research process (from definitions to items in questionnaires) etc. Therefore, bad theory can persist if it is protected by related, (seemingly) consistent, equally bad assumptions that make it seem in line with (seemingly) good evidence (that may actually be based on problematic methodological or epistemological premises).

But can there be such a thing as "good" and "bad theory" in the light of holistic underdetermination? Without entering the philosophical argument on how serious what kind of underdetermination is and how to legitimately respond to it, the history of science teaches us that we should always worry about the "unconceived alternatives" in science (Stanford, 2006) and even the small, useful, but unconceived modifications in existing theories. Still, even if we should never assume that we have already developed the optimal theory in a given area, there are theories we would be more inclined to (temporarily) accept or refute upon systematic reflection in the light of presently existing or new arguments or data. We can indeed conclude that certain theories, for example, rely on ill-defined concepts or contradict cherished assumptions or large amounts of evidence whose methodological basis we do not want to call into question, and those would be considered "bad" theories in comparison to those without such obvious problems.

And, maybe most importantly in the present context, we should be concerned about "bad theory" that is confined to seemingly self-explanatory "falsifiable" hypotheses and that does not make the wider elements and the structure of the network explicit—or at least, does not reflect on them even if large parts of the network cannot be presented in a single publication. In this case, it remains unclear what results mean and how concepts, methods, and findings are related. We cannot be certain about the validity of measurements or the (plausible) causal mechanisms, and may thus also be unaware of potential biases introduced by unsuitable methods and instruments or by neglected aspects.

---

be made in response to contrary evidence (or arguments) (see Stanford, 2021, with reference to classical theorists of underdetermination such as Duhem and Quine).

This can also lead to problematic conclusions. Methods are enshrined or dismissed for the wrong reasons, conceptions and explanations are questioned or left unquestioned based on insufficient arguments. And practical implications (in terms of instrumental usefulness or critical potential) of (supposed) findings may be missed or misrepresented because we lack adequate explanations for our results or because we do not have the right theoretical tools to derive such implications in the first place. Without good theory, we risk to mislead not only ourselves and other researchers, but also the public on what our research means, or to disappoint everyone waiting for our findings to make sense.

Based on a certain philosophy of history or social change, a caricaturesque researcher may be tempted to judge everything either in terms of progress, or, according to their biases, more likely in terms of decline, and hypothesize that the quality of argumentation has decreased with the advent of social media (without explaining whether "quality" is meant normatively or in terms of persuasion). Based on their idea of a universal, essential quality of arguments, they train a classifier to recognize good and bad arguments, using a sample of texts somehow collected online. They then apply this classifier to old newspaper editorials and recent social media comments, ignoring the different functions, stylistic conventions, linguistic features etc. associated with the two genres, and indeed find that comments contain many more bad arguments. The researcher may be convinced to have tested the hypothesis and proudly explain their results to an audience thirsty for this confirmation of their prejudices—that may however, be surprised to learn that arguments must be somewhat better on Instagram than on Twitter, whatever that means (the researcher has no clue).

Merton (1945, p. 462) once summarized the "radical empiricist motto" as follow: "This is demonstrably so, but we cannot indicate its [social and theoretical] significance." I would also add that without a certain theoretical context establishing the validity of the findings, we cannot even convincingly demonstrate that it is so—neither in a single study, nor by "replicating" certain findings. Without a clear conception of what is essential to the phenomenon under analysis and to the methodology of its investigation, it remains unclear what can count as a replication and what to make of seemingly contradictory evidence. Replication should therefore not be defined as the simple repetition of the original procedures, which may only mask a poor theoretical conceptualization of the relevant effects. Only good theory will allow to make a convincing argument that a new study with its specific conditions and its old or new methods, can actually produce new evidence both in favor of, or against an existing

claim (Nossek & Errington, 2020). In the following, I distinguish different more specific problems in theorizing, grouped into a number of types, together with a somewhat catchy terminology that would be suited for critical discussions and reviews of all kinds.

Some Types of Insufficiently Systematic Theorizing

I would call the first interrelated patterns of problematic theorizing "associative asymmetry" and "theoretical cherry-picking." Usually, pairs of constructs are hypothesized to stand in some relationship, and the hypotheses are developed rather associatively. For example, the effect is assumed to mirror the stimulus (e.g., seeing violence leads to violent behavior). Or a type of behavior is explained by a tendency (such as a personality trait) to exhibit behavior from a broader category that includes the type of behavior to be explained (e.g., people with an "aggressive personality"—if we ignore whether this makes sense as a construct—will abuse others online because that is a type of aggression). Or if A has been found to cause B, it may also cause C' that is seen as similar to B (e.g., if reduced revenues of media organizations lead to less diverse coverage in terms of issues, it will also lead to less diversity in terms of actors being covered). Or if A leads to B and B leads to C (at least according to somewhat uncertain earlier research), A will lead to C.

These are of course potentially fruitful ways to arrive at new hypotheses (albeit sometimes rather trivial ones). The problems start when alternatives are not considered systematically ("asymmetry"), and existing theoretical and empirical literature is cited selectively to justify the hypotheses instead of considering a wide range of publications and arguments to arrive at the most plausible hypotheses ("cherry-picking"). While cherry-picking is considered a major issue when it comes to conclusions from empirical findings, it often seems to be considered perfectly acceptable when hypotheses have to be justified (although the literature is usually cited before mentioning the hypothesis, it often seems that it has been searched *ex post* based on the previously established hypothesis).

Systematic theorizing should therefore thoroughly consider arguments in favor of alternative causal orders (e.g., $B \rightarrow A$ instead of $A \rightarrow B$, or C being a mediator instead of a moderator of this relationship), inverse directions of relationships (i.e., a positive versus a negative effect), and alternative forms of relationship (e.g., curvilinear instead of linear). Graphs, cross-tabulation, and other tools can of course help to systematically go

through all the relationships between a set of concepts and to check, modify, and extend models.

Researchers cannot only switch the direction and form of a relationship and check for forgotten factors, but should always consider alternative types of explanations, whether they are immediately tested or only serve to justify a relationship under investigation. Other forms of explanations can also lead to new considerations on the direction and form of relationships and on the inclusion of factors. A number of authors have proposed typologies of explanations—not predictions pertaining to specific observable phenomena or classes thereof, but general "theoretical orientations" that propose certain causal mechanisms involving abstract concepts (e.g., actors who act according to their rational interests or according to social norms, or organizations that tend to legitimize their existence and their control of resources) (see e.g., Bell, 2009; Elster, 2007; Rueschemeyer, 2009; Stinchcombe, 1986).

It may be argued that debatable associations and cherry-picked types, directions, and forms of causal relationships are harmless because non-cherry-picked empirical evidence will weed out the false hypotheses and the potentially right ones can be discovered during data analysis (e.g., when, contrary to the initial hypothesis, a correlation turns out to be negative). However, not all correct relationships will necessarily be identified while testing false hypotheses (e.g., U-shaped forms of relationships in a linear regression or that the unmeasured factor D instead of A is the most important predictor of B). At best, to test the better hypotheses—if they finally come to mind—can require an unnecessary additional round of data analysis of even data collection. Or much space is unnecessarily spent to frame results as surprising and to go on about how they are still inexplicable and how further research is needed to explain them—which will fail unless the new studies are either exploratory and suited to identify the new explanation, or unless such an explanation is finally identified through new efforts of theory building and tested in new studies. At worst, better explanations and interpretations remain undiscovered and untested because data seems to fit the existing ones sufficiently well. Or new empirical research still produces "inexplicable" results because the findings do not speak for themselves, as one may have hoped, and no new explanations have been developed.

Systematic reviews and meta-analyses before or in between empirical studies can help to make sure that one does not cherry-pick from existing research and that the most relevant research gaps and, in particular, the effects that are most in need of further explanation or the most promising theoretical explanations can be addressed (Cooper & Hedges, 1994).

However, this does not prevent cases in which new but *a priori* implausible hypotheses are unnecessarily established and tested on the basis of a one-sided argumentation that is not yet grounded in empirical evidence or not yet made plausible by existing strong theoretical arguments that could be systematically reviewed beforehand.

Asymmetrical theorizing and cherry-picking lead researchers to neglect certain factors or to wrong assumptions about their relationship. In other cases, too many factors are added, also based on problematic assumptions about relationships. I would call these patterns "unstructured listing," and, as a particular subtype, "conventionalized controls." Instead of thinking in terms of processes and causal and temporal order, factors at different levels of abstraction and at different steps of a process are simply added to an unstructured list and to statistical models. In particular, some variables are sometimes only included because they belong to a group of usual control variables (such as age, gender, and formal education).

Assume that in a simultaneous test of the relationships *Attitude A* → *Behavior B* and *Education* → *Behavior B*, we do not find a (strong or significant) influence of education. But in reality, the causal order is *Education* → *A* → *B*, the attitude being the more proximate explanation of the behavior than "education" (or often more correctly, the social background which is approximated by formal education). If one is mainly interested in the attitudinal precursors of some behavior, controlling for education is unnecessary or even dangerous because the influence of A may be underestimated if education is a predictor of the relevant attitudes. But if one were interested in an analysis of socio-demographic or social-structural causes instead of effects of sometimes almost redundant dispositions (B being explained by the tendency to do something like B, what I would call "explanatory triviality"), one would find "education" to be a relevant factor and one should omit A from the theory and analysis. Of course, one can also postulate and test an overall multi-step causal order by means of a path model or it can of course be justified to include "education" as a control variable in order to account for other attitudes that are correlated with both formal education and A but that are difficult to include directly (but education then still remains a potential cause—or a proxy of the causes—of the attitudes). However, such decisions should be made based on good theoretical arguments, not on conventions about what control variables to routinely include. Otherwise, we may come to problematic conclusions, for example that social structure is irrelevant and everything is a matter of attitude, or that almost everything is related to social structure (which is not very surprising and informative in many cases) but that we still not know much about the more specific causes (because the effect

of all more specific causes is "controlled away" by including all kinds of social-structural variables).

A final issue of asymmetric and thus unsystematic thinking is the "fallacy of studying the new to see what's new" instead of systematically comparing it to previous phenomena, both theoretically and empirically, in order to identify actual change. A lot of research on media change and social trends (with all kinds of theses on "-izations") lacks historical depth and appropriate theoretical criteria of comparison with earlier phenomena. This leads to illusions of change based on the wrong levels or dimensions of presumed change or an inadequate picture of a "primitive" past when certain things supposedly did not exist, a "nostalgic" past when current evils did not yet prevail, or a "simple" or "static" past when all the complexity, dynamics, and contradictions that make an analysis of current society challenging were not yet relevant. For example, a supposedly unitary era of the mass media with its corresponding political landscape may easily be idealized as being relatively harmonious, simple to understand, and developing only slowly, as opposed to the turbulence and confusion of the current era.

Problems with the Validity of Theories and Measurements

It has been argued above that the validity of causal inference relies on an adequate theory that helps to specify, among other things, what constructs are to be included, their relationships, the context in which a relationship can be expected, and a design that is consistent with these assumptions. To this question of the validity of causal inference comes the problem of the validity of measurements.

Theorists of validity do not agree on a single conceptualization of validity and the aim of this section is not to provide one but only to point to certain problems that will probably be detrimental to validity under different relevant theories. These problems concern the neglect of theory and interpretation in judgments of validity, or the restriction to "validity by correlation" and the "distant reading of definitions and items," as the problems may be called.

If we accept that to measure means to systematically assign values that we claim to stand in a systematic relationship to something (such as an existing phenomenon or a purpose), validity can be defined as the existence or justifiability of that systematic relationship (the measure is actually caused by that existing phenomenon, see Borsboom, Mellenbergh

& van Herden, 2004, or it is actually justifiable to use the measure for the intended purpose, see Messick, 1989).

In this sense, validity cannot in itself be demonstrated solely statistically but ultimately only interpretively and argumentatively, by relying on a theory about meanings and relationships that can only be tested empirically in parts, if at all.

Unfortunately, certain cues for the validity of measurements have come to more or less replace the originally fruitful and relevant core idea of the concept. For example, the correlation between the construct to be measured and other constructs has been called a type of validity ("criterion validity") instead of being a cue that could alert researchers to certain problems of validity (Messick, 1989).[4]

The problem of validity is then only shifted to the validity of the correlated construct (Messick, 1989). Correlations can vary according to the sample; and a high correlation cannot mean that two measures should be considered as measures of the same thing (otherwise, to the degree that almost everything correlates somewhat with almost everything else, everything would be a measure for everything, albeit a very imperfect one most of the times) (Borsboom, Mellenbergh & van Herden, 2004). The classical proposal by Cronbach and Meehl (1955) also does not go far enough. They assume that potentially complex "nomological nets" are built and rebuilt over time that connect constructs and tie them to observations and thus ensure the validity of a construct in question. They also emphasize that these networks of relationships have to be theoretically interpreted. However, the main issue for the validity of constructs cannot be to establish a network of theoretically interpreted but most importantly, empirically substantiated relationships, but a network of relationships made of assumptions about how the measurement can refer to what is claimed to be measured, and how the measurement is produced (thus, 1. a theory of meaning as reference to an existing phenomenon and its pro-

4 Of course, it is possible to define "validity" as on wishes to. However, if statistical tests dominate the evaluation of measurements (regardless of whether they are considered an assessment of "validity"), an important aspect of this evaluation would be lost. For example, overviews on the concept of validity aimed at communication researchers classify validity into different forms, such as "construct," "content," "face," "convergent," "discriminant" etc. and do in fact mention the role of theory or that an *argument* has to be made for the validity of a measurement (see, e.g., Dilbeck, 2017; Fink, 2017; Martinez, 2017). However, the logical relationship between what is mostly called "types" of validity (instead of strictly considering it as cues for validity) and validity proper is not always clear and statistical cues feature more or less prominently in such overviews.

perties and 2., in the case of questionnaire-based instruments, a theory of response behavior, see Borsboom, Mellenbergh & van Herden, 2004). Or if we do not subscribe to a realist ontology and theory of meaning in which a measurement refers to something that actually exists in the most direct sense of the word (which would then cause the measurement, Borsboom, Mellenbergh & van Herden, 2004), we need a theory of meaning that ties the measurement to a description of what it is supposed to do or represent, and an explanation of how it can achieve that.

Here, we are again faced with a challenge of holism: We need a theoretical network of interconnected definitions and semantic, causal, and other relationships that we consider consistent and—as far as some of the parts involve truth claims—to be true (but individual assumptions again cannot be falsified in isolation. In particular, a weak or counter-intuitive correlation cannot decide about the validity of a construct but must be considered in the context of the whole network). This network makes sure that we can systematically connect to measurement to its meaning.

Apart from the other theoretical considerations involved in the assessment of validity, I would like to emphasize the role of interpretation, in particular the careful and informed reading of definitions and of questionnaires (if we restrict ourselves to standardized interviews as an example). I often have the impression that the interpretive work and the work of logical deduction and argumentation in the context of measurement is not as careful as it could be (even if it concerns definitions established by the researchers themselves). For a measurement to be valid, we first have to ask ourselves whether we really include all aspects and nothing else than what is covered by the definition of a concept—which is of course a standard requirement for validity. However, this means to carefully apply the criteria in the definition to different candidate cases. Furthermore, we have to ask whether questions and items mean what we think they mean in the ordinary language of all relevant social groups that are to be included in a study. This requires interpretive skills, a particular sense for everyday language, or almost ethnographic experience and knowledge, as well as argumentative and logical rigor, and systematic "testing" of the scope of concepts (by discussing whether diverse and systematically selected examples fall into the definition and whether this is intended or fruitful. We should thus ask: "Can we reliably decide whether this is included in the definition?" and: "Do you really want that to be included in your definition?"). Only based on careful and well-substantiated interpretations, we can then disentangle the whole network of theoretical of assumptions that is supposed to guarantee the validity of some measurement.

For example, we have to assume, argue, or empirically demonstrate that ordinary speakers of the English language would interpret the word "politicians" in the statement "I trust politicians to work in the interest of ordinary people" to refer to political actors at all levels of government, from local to supranational, if that is our definition of "politicians" and our measurement of trust of politicians in a questionnaire. One may well doubt that respondents mostly think about mayors and EU Commissioners when they read this statement. Ultimately, if the validity of this measurement in relation to the above definition is questioned, what counts will be arguments or evidence on the typical interpretation of the word "politicians" (and not in general but if used in a statement such as the above). Maybe, the definition of "politician" may also turn out to be problematic—for example whether "government" refers to "government" as in "all branches of government" or "government" as the executive branch (whatever that means at the local level, depending on the system of "government"). And these are rather simple questions compared to the ontology of trust and its potential objects, and to a theory whether and how it can actually manifest itself in responses to such an item. Therefore, the more atheoretical and methodologically or epistemologically less elaborate among the studies on changes in political or media trust should be taken with a grain of salt, in particular as the interpretation of concepts and measures does not only refer to a single point in time.

If only "face validity" (e.g., the items make sense and one simply hopes that everyone will agree on the meaning and relation to the construct, and once a measurement is established, its meaning seems self-evident to those working in the field and it is no longer questioned) and statistical tests for convergent and discriminant validity are required, it is relatively easy to establish new constructs and easy for them and their measurements to persist. We are then subject to the "dictate of cumulativity" and can walk into a "reification trap."

We witness a trend toward standardization in communication research, from the canonization of methods (e.g., in introductory textbooks) to well-documented and reusable scales and other measurements. One of Ioannidis' (2005, p. 698) corollaries states that "flexibility increases the potential for transforming what would be 'negative' results into 'positive" results'" and that "adherence to common standards is likely to increase the proportion of true findings." According to his explanations of this corollary, he seems to have thought of two factors affecting the validity of measurements and data analyses and thus of the results: new versus time-tested methods and room for selectivity and manipulation. However, his reasoning seems to be biased toward standardization—as many communication

researchers obviously seem to value fixed methods and measurements, and established constructs in general, for different reasons. The general belief is that science progresses if studies with a comparable basis accumulate.

Theories (!) of standardization suggest that the reduction in complexity, the gains in compatibility or comparability can come at the price of problematic lock-ins: A norm is perpetuated not because it is the best solution but because a break would come at certain costs (in the case of research, data can no longer be fused and time series cannot be extended, reviewers may reject divergent methodologies, etc.; see, e.g., David, 1985, for a famous explanation of lock-ins due to technological standardization which has also been subsequently applied to institutional path-dependencies).

In many fields, a convention is all that is needed to fulfill the functions and realize the gains of standardization. It is often more important that a standard exists than what the actual standard is, as long as it is in the range of sufficiently functional alternatives. However, if we believe that certain methods and measurements are superior to others, a well-justified choice cannot be replaced by convention, and has to be grounded in substantial conceptions of the object of study and the procedure. And we did not even enter the discussion on paradigm shifts and similar breaks that, according to different theories of science, lead to progress or new incommensurable but acceptable perspectives. Anyway, the idea that research is additive and progresses as long as its building blocks (new studies or new elements of theories that do not change the whole) are compatible, is a rather strong assumption both globally and with regard to specific objects of study, and the requirement that when in doubt, they should remain compatible (the dictate of cumulativity) has to be questioned in each individual case.

In terms of methodological and substantial theory, the opposite problem of "reinventing the wheel" can of course also be observed, and the systematic review and use of existing theories, concepts, and measurements is a potential solution. As another remedy, we should routinely and systematically identify superordinate categories, functional equivalents, or otherwise similar phenomena to the ones under investigation and check whether there are already theories, findings, and measurements pertaining to them. There is no need to theorize or operationalize a phenomenon *ad hoc* if we already have a convincing more general theory that applies to it, and we do not have to start from scratch if similar phenomena have already been theorized and investigated, risking to fall back behind existing approaches. Research do not have to repeat the same mistakes of schematic *ad hoc* theorizing for each new media technology, genre of content, application, organizational innovation, or social trend.

If we uncritically stick to existing conceptualizations or keep postulating new concepts *en passant*, and do not reflect on the broader context of our concepts and the underlying assumptions, we risk to commit ourselves to messy and contradictory ontologies or discredited epistemologies—which puts the validity of empirical findings into question on a much more fundamental level. Not only can there be a simple mismatch between some concept and some measurement, but we risk potentially fundamental category mistakes, for example by confusing statements on meaning with statements of facts (*A means B for actor C* with *A is B*), the perspective of observers with that of actors, or normative with factual claims.

For example, if we were to define "disinformation" as statements or sets of statements that we know to be false and that the communicator knows to be false, what does it mean for someone to be exposed to disinformation? The only thing that person is exposed to is the statements, so research on consequences of disinformation has to discuss whether "disinformation" can really be category of reception and effects research, because some of the defining features, such as the knowledge and intention of the communicator, are not really present in the situation of reception. It is easy to propose a number of hypotheses using the concept of "disinformation," either *ad hoc* or based on a number of known principles of persuasion, but it is important to reflect on the ontological and epistemological foundations of research that involves multiple perspectives on the truth of statements. Otherwise, the scope, meaning, and implications of empirical findings remain unclear: What kinds of statements do the results apply to, what were the actual mechanisms of persuasion (or resistance), and what kind of competence would recipients need in the present information environment or in the light of potential new types of disinformation (Krämer, 2021)?

In sum, it is problematic to postulate concepts by virtue of an existing measurement with certain statistical properties and to perpetuate them by virtue of their existence in the literature and of the continued use of the measurement. We should not unreflectedly reify concepts and uncritically "blackbox" constructs by routinely applying some operationalization without discussing the underlying assumptions and processes.

One aspect of such uncritical postulates of conceptualizations is what has been termed the "scholastic fallacy": "To place the models that the scientist must construct to account for practices into the consciousness of agents, to operate as if the constructions that the scientist must produce to understand practices, to account for them, were the main determinants, the actual cause of practices" (Bourdieu, 1990, p. 384). If we ignore the difference between the logic of "theoretical practice" and everyday "practical

practice," we can establish all kinds of explanations and operationalizations that project a model into some subjects without checking whether it actually grasps their thinking and doing. Furthermore, the scholastic fallacy tends to take concepts outside the context where they are adequate and to assume that a given theory or the perspective of researchers from a given epoch, social class etc. are universally valid.

Carleheden (2016) criticizes the rather naive assumptions of a protagonist of the discourse on systematic theorizing, Richard Swedberg, who seems to imply that theorizing becomes more realistic and less out of touch if efforts of theorizing are preceded by some observations in the field. Such quick and dirty pilot studies, if done unsystematically, cannot be very fruitful (Tavory, 2016) and theorizing cannot be based on supposedly atheoretical creative and open-minded empirical research that is later turned into more formalized hypotheses (Carleheden, 2016).

However, it still seems important to find a third alternative to the naive realism of a belief in pre-theoretical data or experience, and the sterile modeling and operationalizing without any close contact to a field or corpus. The alternative would be theoretically informed and systematic but open qualitative research that reflects and continually adapts its theoretical basis, methodology, sampling etc., and systematic theorizing that critically draws on existing research and is accompanied by an ongoing engagement with the corresponding social fields. Often, this not only means to be present on the latest platforms, follow the latest trends, and talk to people about them instead of only plugging together constructs, but also to bury oneself in old newspaper articles, interviews, or other sources to really get a sense of past political, popular, or intellectual culture.

Problems with Critical Reflection of Studies and with Conclusions

Beyond the explanations a study can offer (often narrowly referred to as *the* "theoretical contribution" it makes), theory should also inform methodological reflection. This can range from a few simple thoughts on the behavior of (potential) participants (as in the example in the introduction) to fundamental methodological questions. Otherwise, concluding section of publications are often restricted to what I would call "ritualistic limitations" instead of actual reflection. The usual restrictions of the respective type of study are reiterated because it is custom to do so or maybe because one wants to preempt obvious criticism: cross-sectional designs do not allow for causal inferences based on temporal order, a convenience sample is not representative, self-reports may be biased by social desirability, etc.

More elaborate conceptions of the design and measurement process and of potential biases can render this discussion much more informative and instructive: What is the process that generates the data and are there any biases (in relation to what exactly?), does it apply to other phenomena and methods as well, is it possible to argue for a specific type and direction of bias, etc.?

Theory cannot only provide the framework for empirical research but if theory is defined as more than testable hypotheses, it can also serve as an interpretive framework or guide through *praxis* and through society. However, due to a restrictive concept of theory and the lack of familiarity with theories of society and normative theories, scholarly works in communication research often suffer from what may be called "diagnostic" and "normative triviality."

If the social sciences wish to offer more than isolated findings (some of which may of course be highly critical in a given historical situation), they should also embrace their function of making sense of the times and world we are living in (*Zeitdiagnose*, i.e., diagnosis of the times; Junge, 2016) and not leave this task to other commentators (who are of course entitled to their judgments and to participate in open debates, but who may be less familiar with current research and less skilled in conceptualizing certain phenomena).

Scholarship can then offer meaning or concepts to think or argue with in the public sphere and in conversations outside the scientific field instead of only taking up the debates and buzzwords of the day. Often, certain concepts of metaphors shape public discourse much more strongly than empirical findings as such that may or may not be in line with those concepts (think of "filter bubbles," "cancel culture," "information society," or "fake news"). While researchers should of course make sure that the ideas they offer to public discourse are not in direct contradiction to empirical research or other established standards of evaluation, such concepts are not necessarily just repackaged or (over-)generalized results, but more often schemata that group similar phenomena or descriptions and that highlight certain features of a phenomenon or a whole era, often with a critical tone.

C. Wright Mills emphasized the role of sociology—or, one might say, the social sciences in the broadest sense—to offer orientation beyond mere factual information (while also describing his "craftsmanship" and how he thinks "sociological imagination" can be methodologically stimulated, providing early but still useful insight into strategies of theorizing and conceptualization and the development of research interests):

> "The very shaping of history now outpaces the ability of people to orient themselves in accordance with cherished values. [...] Is it any wonder that ordinary people feel they cannot cope with the larger worlds with which they are so suddenly confronted? That they cannot understand the meaning of their epoch for their own lives? […] It is not only information that they need – in this Age of Fact, information often dominates their attention and overwhelms their capacities to assimilate it. [...] What they need, and what they feel they need, is a quality of mind that will help them to use information and to develop reason in order to achieve lucid summations of what is going on in the world and of what may be happening within themselves. It is this quality, I am going to contend, that journalists and scholars, artists and publics, scientists and editors are coming to expect of what may be called the sociological imagination."
>
> (Mills, 1959, p. 4f.)

One might object that today, we do not live in an "Age of Fact" but an age of disinformation—but is that the case and what would be the theory that would provide clear criteria to decide? Be that as it may—if information from generally trustworthy sources is available more abundantly and more easily than ever, can even more information (in particular coming from researchers) be the solution? Certainly, specific kinds of knowledge can and should always be made more accessible to the public. But will it be understood and trusted without more general frames of interpretation that help us to make sense of social relationships and of social fields such as science, the media, or politics, and society?

Of course, such interpretations should not be thought of as authoritative guidelines conveyed by scientific eminences, but something that everyone should be able to challenge and adapt in open discourses. However, facticity in public discourse is not only about single statements and small pieces of evidence, but also about well-justified general frames of interpretations and worldviews in which individual claims about reality do or do not make sense and do or do not appear plausible or correct.

To offer this kind of interpretations, and to offer better ones than those already circulating, does not only require a certain amount of creativity, but also the readiness to transgress the boundaries of a scientistic habitus, i.e., a deep-seated sense of what is good and "real" science: only the most rigorous empirical research based on the most arid terminology, avoiding anything that could come close to speculation or editorializing.

While many empirically oriented researchers will be rather unfamiliar with, but aware of theories of society and more abstract and interpretive

social theory, they probably are even more "unmusical" when it comes to normative theories (to start with a catchy metaphor, which is one of many techniques in theory building!). Almost anyone can hum a simple melody they have heard or invented, and many publications in communication research mention at least some critical or normative implications. However, many publications and also many personal conversations seem to suggest that normative statements are either equated with references to commonsense social problems and dominant norms (such as professional norms in some field or basic norms of liberal democracy), or they are lumped together with personal "opinion" or ideology, and thus something "subjective" that is to be avoided in serious research—something like expressing one's musical taste, which can be interesting, but nothing interlocutors will be able to agree on based on the better arguments. As a piece of music (whether it is composed or even improvised) is not a random invention or intuition but something that relates its elements following or ostentatiously breaking certain rules, a normative theory is also a structured whole with basic concepts, logical or argumentative relations, presuppositions and implications, criteria of consistency or contradiction and tensions, etc.

As in all processes of theorizing based on metaphors, we have to decide how far we are willing to follow them. In the present case, it is questionable whether "good" music and "good" theory can be fruitfully compared. Be that as it may, normative and other interpretations, conclusions, and contextualization of research findings should strive for the same argumentative rigor and systematicity as empirical research or the development of those more specific statements that immediately guide data collection and analysis.

For example, a lot of interest in media change is ultimately motivated by the question of whether new developments have brought progress or decline (even if those emphatic terms themselves are rarely used) with regard to democracy, health or wellbeing, equality etc. However, what can count as positive and negative developments is often left to commonsense and rarely explicated in terms of a consistent and well-justified normative theory. It is taken for granted that readers will agree that increasing "fragmentation" is a bad thing, that people should assume responsibility for their health or the environment and should be persuaded to do so by the most effective messages in new media formats, that high levels of trust in the institutions and actors of liberal democracy are desirable and social media companies should help to achieve this goal, or that science communication should more than ever be based on randomized studies, meta-analysis, and similar types of "high quality" evidence. However, I would suggest that in each of these cases, elaborate normative reasoning

would lead us to answer with a "Yes, but…" or a "It depends what is exactly is meant by…"

Conclusion

Bad theory can lead to a waste of resources, biased analyses, and to a lack of understanding of an object of study and the social world, so that critical and practical implications of our research remain undiscussed and we can neither offer sound evidence nor interpretations and diagnoses to the public. Good theory leads to focused research efforts, appropriate methodologies, and valid measurements and analyses, and allows us to offer the public not only isolated findings, but concepts to think and discuss with and to understand society. It leads us out of an interpretation crisis that does not only concern individual results but also calls into question whether larger fields of research can really contribute to our understanding of our current social world and era. We should not restrict ourselves to the replicability of certain methodological aspects while testing isolated hypotheses (as important as this is), but aim for the transparency and critical reflection of interpretations and theoretical assumptions in the broadest sense. Without this attention to theory, it does not make much sense to retrace, repeat, and accumulate research—no replication without interpretation, no validity without theory, but also no true originality without reality checks. The first step toward better theory is to recognize potential causes for shortcomings, such as unsystematic theorizing and the lack of fruitful interpretive frameworks, resulting in questionable validity and a superficial reflection of implications.

Systematic theory building is thus a necessity in all projects, not a hobby of a few thinkers or a closed field that is separate from empirical research. Of course, a certain specialization is inevitable and functional, and methodological experts can always collaborate with good theorists, but they also need certain theoretical knowledge and competences in order to reflect their work beyond the methodological technicalities and to identify points of contact when working together.

When it comes to competences of theory building, I would bet that most graduates in communication have read a book or attended a course on methods of data collection, but most of them have never looked into a book on theorizing or even taken a course on the subject (i.e., courses on theorizing, not merely courses on theory), and might even rarely read explicitly theoretical contributions in the strict sense.

Still, one should not stand in awe of theorizing and see it as an irrational process that only geniuses master intuitively. Learning to theorize requires two kinds of practice:

1. reading actual theory, in particular genuinely theoretical contributions, and paying close attention on the theoretical sections of the theoretically most sophisticated empirical publications (with many concepts and approaches thus stored somewhere in long-term memory, we can also partly rely on intuitive theorizing, i.e. establish associations through heuristic processing even if we are not focusing on the specific topic of research or on the task of systematic theory building, see Knorr-Cetina, 2014), and
2. practicing theory building based on publications that codify the process, that specifically turn to the "context of discovery" (Swedberg, 2012), focusing on theorizing as opposed to theory—although existing overviews are often rather restrictive or idiosyncratic in their understanding of theory, and the procedures they propose remain rather abstract. This more general instruction therefore needs to be complemented by teaching or collaborative learning based on specific examples and a broader range of problems. This also allows theorizing to become more intuitive and maybe less painstaking over time, although we should always check its results for its systematicity (avoiding, for example, asymmetries or category mistakes).

The idea of systematic theorizing has been met with enthusiasm, sobering qualifications, pragmatism, and constructive criticism. After an initial optimism in the 1950s and 60s that the main social "laws" might soon be discovered, and the more modest proposal of grounded theory to build theory inductively from empirical material, a newer "pragmatic" wave of literature seeks to stimulate theory building based on heuristics and tricks for creativity (Tavory, 2016). However, this newer approach has also been criticized as overly narrow: It tends reduce the necessarily cyclical interplay between theorizing and empirical research to a strict distinction between the context of discovery and the context of justification that is reproduced uncritically, and it tends to reduce theory to causal explanations (Büttner, 2021; Tavory, 2016).

German-speaking scholars in particular seem to have increasingly turned toward different methods and aspects of building social theory that go beyond a narrow conception of theory as set of falsifiable propositions (e.g., Anicker, 2020; Büttner, 2021; Beregow, 2021; Farzin & Laux, 2014; Farzin & Laux, 2016; Krämer, 2015). Furthermore, the practice of doing theory has to be reflected as a social activity (not only the activity of an individual genius or craftsperson) that needs a sound basis in experience and existing thought, but that is not automatically without biases because it is done systematically, and that always requires a critically distance to

previous approaches. We need to be aware of how contexts shape the production and circulation of theory and how doing theory always risks to shift from making, teaching, and critical analysis of theories to the consumption of ready-made, hegemonic, commonsense ideas (Chaudhuri & Thakur, 2018).

Unfortunately, teaching on theorizing is still often limited to some basics of Popperian philosophy of science, a few remarks on criteria for good theory (consistency, falsifiability, etc.), or some ways of coding qualitative material in methodology courses and textbooks. The "imbalance between methods and theory" (Swedberg, 2014a, p. 8) and between teaching theory and teaching theorizing needs to be overcome by more explicit teaching of the practice of theory building (Silver, 2019; Swedberg, 2014a, 2014b, ch. 7; Swedberg, 2016) at all levels of scientific qualification. It should be based on literature that conveys the methods of theorizing that is not restricted to specific types of theories or particular problems and aspects of theory building, and that includes the critical reflection on the contexts and biases of theorizing.

I think that it would also be a good idea to remember a number of teachings by a researcher who, unlike Ioannidis, did not put his approach to science in writing, let alone in such pithy words. Wolfram Peiser, whom the present volume is dedicated to, always advised me and his other students and collaborators to consider a broad range of explanations and factors, to systematically think about inverse relationships, and only thereafter focus on a range of concepts to be included in a study (and when in doubt, to include more instead of fewer as long as everything can be measured parsimoniously but validly and reliably). Instead of focusing on the next best idea, he always reminded everyone to think broadly, choose wisely, and explain their choices. He also urged everyone not to be narrow-minded due to a single preferred and closed theoretical framework but to check for alternatives and possible additions, to use common sense in order to find the most fruitful questions and adequate explanations, and to connect one's research to broader concerns and debates inside and outside the scientific field. Finally, he insisted that what is specific to a phenomenon or what is new can only be identified by systematic comparison. In particular, he always reminded us that claims of social or media change must be based on a systematic comparison with the past (whether based on original or existing research and data), not only on the study of the most recent phenomena. In this sense, much of the present argument is not that new but inspired by what he taught us.

References

Abbott, Andrew (2004). *Methods of discovery. Heuristics for the social sciences*. New York: W. W. Norton.

Abend, Gabriel (2008). The meaning of 'theory'. *Sociological Theory, 26*, 173-199.

Alexander, Jeffrey C. (1982). *Positivism, presuppositions, and current controversies [Theoretical logic in sociology, vol. 1]*. Berkeley: University of California Press.

Anicker, Fabian (2020). Theoriekonstruktion durch Theorienvergleich – eine soziologische Theorietechnik. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 72*, 567-596.

Becker, Howard S. (1998). *Tricks of the trade. How to think about your research while you're doing it*. Chicago: University of Chicago Press.

Bell, David C. (2009). *Constructing social theory*. Lanham: Rowman & Littlefield.

Bendavid, E., Mulaney, B., Sood, N., Shah, S., Bromley-Dulfano, R., Lai, C., ... & Bhattacharya, J. (2021). COVID-19 antibody seroprevalence in Santa Clara County, California. *International Journal of Epidemiology, 50*, 410-419.

Beregow, Elena (2021). Theorieatmosphären. Soziologische Denkstile als affektive Praxis. *Berliner Journal für Soziologie, 31*, 189-217.

Borsboom, Denny, Mellenbergh, Gideon J. & van Herden, Jaap (2004). The concept of validity. *Psychological Review, 4*, 1061-1071.

Bourdieu, Pierre (1990). The scholastic point of view. *Cultural Anthropology, 5*, 380-391.

Büttner, Alexander (2021). Theoretisieren in der Soziologie. Bezugskontexte und Modi soziologischer Theoriebildung. Available online: https://doi.org/10.31235/osf.io/bj42k

Carleheden, Mikael (2016). What conception of the theoretical does 'theorizing' presuppose? Comment on Richard Swedberg's 'Before theory comes theorizing or how to make social science more interesting.' *British Journal of Sociology, 67*(1), 36-42.

Chaudhuri, Maitrayee & Thakur, Manish (Eds.) (2018). *Doing theory. Locations, hierarchies and disjunctions*. Hyderabad: Orient Black Swan.

Cooper, Harris & Hedges, Larry V. (1994). Using research syntheses to plan future research. In Cooper, Harris & Hedges, Larry V. (Eds.), *The handbook of research synthesis* (pp. 485-500). New York: Russell Sage Foundation.

Cronbach, Lee J. & Meehl, Paul E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

David, Paul A. (1985). Clio and the Economics of QWERTY. *American Economic Review, 75*, 332-337.

Dienlin, Tobias, Johannes, Niklas, Bowman, Nicholas D., Masur, Philipp K., Engesser, Sven, Kümpel, Anna S., ... & De Vreese, Claes (2021). An agenda for open science in communication. *Journal of Communication, 71*, 1-26.

Dilbeck, Keith A. (2017). Validity, construct. In Allen: Mike (Eds.), *The SAGE Encyclopedia of Communication Research methods (vol. 4)* (pp. 1820-1822). Thousand Oaks: Sage.

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

Elster, John (2007). *Explaining social behavior. More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.

Farzin, Sina & Laux, Henning (2016). Gründungsszenen – Eröffnungszüge des Theoretisierens am Beispiel von Heinrich Popitz' Machtsoziologie. *Zeitschrift für Soziologie, 45*, 241-260.

Farzin, Sina & Laux, Henning (Eds.) (2014). *Gründungsszenen soziologischer Theorie*. Wiesbaden: Springer VS.

Fink, Edward L. (2017). Validity, measurement of. In Allen: Mike (Eds.), *The SAGE Encyclopedia of Communication Research methods (vol. 4)* (pp. 1828-1833). Thousand Oaks: Sage.

Gigerenzer, Gerd (2010). Personal Reflections on Theory and Psychology. *Theory & Psychology, 20*(6), 733–743.

Hagen, Sebastian, Frey, Felix & Koch, Sebastian (2015). Theoriebildung in der Kommunikationswissenschaft. Eine Bestandsaufnahme zu Bedeutung, Arten und Verfahren der Theorieentwicklung. *Publizistik, 60*, 123-146.

Hoyningen-Huene, Paul (1987). Context of discovery and context of justification. *Studies in History and Philosophy of Science, 18*, 501-515.

Ioannidis, John P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2(8)*, e124.

Jaccard, James & Jacoby, Jacob (2020). *Theory construction and model-building skills. A practical guide for social scientists [2nd ed.]*. New York: Guilford Press.

Junge, Matthias (2016). Zeitdiagnose als Chance der Soziologie. In Junge, Matthias (Eds.), *Metaphern soziologischer Zeitdiagnosen* (pp. 51-60). Wiesbaden: Springer VS.

Kalthoff, Herbert, Hirschauer, Stefan, Lindemann, Gesa (Eds.) (2008). *Theoretische Empirie. Zur Relevanz qualitativer Forschung*. Frankfurt a. M.: Suhrkamp.

Knorr-Cetina, Karin (2014). Intuitionist theorizing. In Swedberg, Richard (Eds.), *Theorizing in social science. The context of discovery* (pp. 29-60). Redwood City: Stanford University Press.

Krämer, Benjamin (2015). Luhmanns Operationen: Analyse einiger Wendungen der soziologischen Theoriebildung. Available online: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-438008

Krämer, Benjamin (2021). Stop studying "fake news" (we can still fight against disinformation in the media). *Studies in Communication and Media, 10*, 6-30.

Krause, Monika (2016). The meanings of theorizing. *British Journal of Sociology, 67*, 23-29.

Lee, Stephanie M. (2020). JetBlue's founder helped fund a Stanford study that said the Coronavirus wasn't that deadly. Buzzfeed News. Available online: https://www.buzzfeednews.com/article/stephaniemlee/stanford-coronavirus-neeleman-ioannidis-whistleblower

Martinez, Lourdes S. (2017). Validity, face and content. In Allen: Mike (Eds.), *The SAGE Encyclopedia of Communication Research methods (vol. 4)* (pp. 1823-1824). Thousand Oaks: Sage.

Merton, Robert K. (1945). Sociological theory. *American Journal of Sociology, 60*, 462-473.

Messick, Samuel (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*, 5-11.

Mills, C. Wright (1959). *The sociological imagination*. Oxford: Oxford University Press.

Nossek, Brian A. & Errington, Timothy A. (2020). What is replication? *PLOS Biology, 18(3)*, e3000691.

Oberauer, Klaus & Lewandowsky, Stephan (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*(5), 1596-1618.

Popper, Karl (1935). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Springer.

Rains, Stephen A., Levine, Timothy R. & Weber, Rene (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association, 42*, 105-124.

Rueschemeyer, Dietrich (2009). *Usable theory. Analytic tools for social and political research*. Princeton: Princeton University Press.

Runciman, W. G. (1983). *A treatise on social theory. Vol. I: The methodology of social theory*. Cambridge: Cambridge University Press.

Shoemaker, Pamela J., Tankard, James William Jr. & Lasorsa, Dominic L. (2004). *How to build social science theories*. Thousand Oaks: Sage.

Silver, Daniel (2019). Theorizing is a practice, you can teach it. *Canadian Review of Sociology, 56*, 130-133.

Sohlberg, Peter & Leiulfsrud, Håkon (Eds.) (2017). *Theory in action. Theoretical constructionism*. Leiden: Brill.

Stanford, Kyle (2021). Underdetermination of scientific theory. In Zalta, Edward N. (Eds.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/win2021/entries/scientific-underdetermination/

Stanford, Kyle S. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. New York: Oxford University Press.

Stinchcombe, Arthur L. (1986). *Constructing social theories*. Chicago: University of Chicago Press.

Swedberg, Richard (2012). Theorizing in sociology and social science: Turning to the context of discovery. *Theory and Society, 41*, 1-40.

Swedberg, Richard (2014a). How to teach theorizing: Exercises and more. *Perspectives. Newsletter of the ASA Theory Section, 36(1)*, 1-10.

Swedberg, Richard (2014b). *The art of social theory*. Princeton: Princeton University Press.

Swedberg, Richard (2016). Reply to the commentators. *British Journal of Sociology, 67*, 57-70.

Tavory, Iddo (2016). The pragmatist wave of theory construction. *British Journal of Sociology, 67*, 50-56.

Benjamin Krämer *(Dr. rer. pol., LMU Munich, 2012) is Akademischer Rat auf Zeit and Privatdozent at LMU Munich's Department of Media Communication. His research focuses on political communication (with an emphasis on populist communication), media use and reception, and online communication as well as on the development and adaptation of new empirical methods and strategies of systematic theory building. Wolfram Peiser supervised his PhD and was a member of the supervision committee of his habilitation in Munich. Benjamin worked as research and teaching associate at his chair from 2006 to 2021.*