Process tracing in policy research

Peter Starke

Contents

Summary

Process tracing is a research method for causal explanation in which diverse empirical observations within one or more cases are understood as implications - or "traces" - of theoretical causal mechanisms. The most complete possible empirical reconstruction of causal processes through case studies allows conclusions to be drawn about (alternative) theoretical explanations. The article the most important features of the method, examples from policy research, strengths, weaknesses and methodological quality criteria.

Keyword¨rter

Qualitative methods - research design - case studies - research method - case selection

P. Starke ( )✉
University of Southern Denmark, Odense, Denmark
E-mail: starke@sam.sdu.dk

G. Wenzelburger, R. Zohlnhöfer (eds.), *Handbuch Policy-Forschung*,
https://doi.org/10.1007/978-3-658-05678-0_17-1

1 Introduction: What is process tracing?[1]

Process tracing - a German translation such as "process analysis" has not become established in research - is one of the most important research methods for qualitative case studies in political science and policy research. On the one hand, it differs as a method for causal explanations from qualitative discourse analytical and hermeneutic methods - despite more recent approaches (Nullmeier 2019; Norman 2015). On the other hand, in contrast to many primarily case-comparative and/or quantitative methods, the focus is not on the correlations between independent variables and the dependent variable, but on the causal mechanisms that link the independent and dependent variables and that can be proven within one or more cases. This can serve either the theory test, theory generation or the explanation of a specific case. The number of cases - "N" in the sense of units of investigation - is usually small in process tracing, but the data basis - "N" in the sense of empirical observations - is very broad and diverse. A wide variety of data collection methods and data formats can be used for process tracing, as can secondary data. In this respect, the method is similar to historical studies, but has a much more theoretical, abstract approach. As will be explained below, the validation of theoretical hypotheses in process tracing is achieved through case- and context-specific conclusions. Empirical observations are understood as potential empirical implications - or "traces" - of a theory. This tracing follows specific rules that help to prioritise and correlate case observations. To summarise, *the definition* on which this article is based is as follows: Process tracing is a research method for causal explanation in which diverse empirical observations within one or more cases are understood as potential implications of theoretical causal mechanisms. The most complete possible empirical reconstruction of causal processes through case studies allows conclusions to be drawn about (alternative) theoretical explanations.

A number of approaches are relatively closely related to process tracing, including "analytic narratives" (Bates et al. 1998; Levi 2002), "causal reconstruction" (Mayntz 2002), "systematic process analysis" (Hall 2006) and "causal narrative" (Mahoney 2003). For others, the connection is rather loose: "pattern matching" (Campbell 1975), for example, is based on some similar elements - such as the attempt to prove implications of theories within one or a few cases - but causal mechanisms are of less importance than in process tracing. Even more clearly, process tracing must be distinguished from older "small-N" methods such as Mill's methods (1968 [1872]), as the latter do not recognise the systematic differences in the characteristics at the

[1]I would like to thank the editors and the participants of the authors' workshop especially Claudius Wagemann, Detlef Jahn, Diana Panke and Jale Tosun. I am also grateful to Henning Deters, Ingo Rohlfing and Frank Stengel for their detailed comments.

independent and dependent variables that form the basis for causal inferences several cases (*cross-case analysis*) (critically, Lieberson 1991).[2]The same applies to the "*congruence method*" (George and Bennett 2005, pp. 181-204; Beach and Pedersen 2016). Process tracing, on the other hand, is based on patterns *within* the analysed case or cases (*within-case analysis*). At the same time, the *research method* process tracing is not synonymous with the *research design* (individual) case study or case comparison.[3]The association of process tracing and case study designs is based more on pragmatic research decisions than on methodological necessities.

This article aims to define the concept and practice of process tracing more precisely and to illustrate its possibilities (and limitations) using a series of examples. For further reading, we refer in particular to the programmatic monograph by George and Bennett (2005), an edited volume by Bennett and Checkel (2015a) and the textbooks by Rohlfing (2012) and Beach and Pedersen (2012). There are also various short introductions in the form of articles (Collier 2011; Ricks and Liu 2018; Checkel 2008; Trampusch and Palier 2016). A monograph by Nullmeier has also recently been published in German (Nullmeier 2021).

The article is structured as follows: A brief historical outline is followed by an explanation of the theoretical assumptions on which process tracing is based and the research objectives that are usually pursued with it. The selection of cases is closely related to this. I then turn to the data that typically used for process tracing. Data analysis, in particular the weighting of observations and the central types of conclusions are the subject of the following section. This is followed by a small selection of examples from policy research. The strengths, weaknesses and quality criteria of process tracing are discussed in the penultimate section. The conclusion summarises the article.

2 Origin, assumptions, research objectives and case selection

2.1 The development of the method

The term process tracing originated in psychology more than 40 years ago and originally referred to a method of analysing cognitive decision-making processes on the basis of direct statements made by the test subjects, observable behaviour

[2]The difference method and the concordance method are the best-known methods described John Stuart Mill. Originally intended exclusively for experimental research, they had a significant influence on so-called "small-N comparisons" in the social sciences (Skocpol 1979). In political science, the "*most similar systems design*" by Przworski and Teune (1970) is characterised by a great similarity to Mill's difference method.

[3]At the same time, it must be distinguished from survey *methods* - such as semi-structured interviews - which can also be used for other research methods.

as well as physiological and neuropsychological data (Ford et al. 1989; Einhorn et al. 1979; Hogarth 1974; Kühberger et al. 2011, on the history
see Nullmeier 2021 for details). In the late 1970s, the political scientist Alexander L. George adopted the method and developed it further, initially on his own (1979) and later together with Timothy McKeown (George and McKeown 1985). In contrast to the psychological literature, however, not only cognitive but also social processes were to be included in the analysis (see Bennett 2008a; Bennett and Checkel 2015b). However, it took another 20 years for process tracing to reach the political science mainstream. The book *Case Studies and Theory Development in the Social Sciences* by Alexander George and Andrew Bennett (2005) was the decisive impetus for this. Process tracing has since become part of the canon of methods used in political science research. Today, the method is far less threatened by marginalisation than by a "buzzword problem […], where process tracing is mentioned, but often with little thought or explication of how it works in practice" (Bennett and Checkel 2015b, p. 4). The aim of this article is also to show that process tracing is more than just a buzzword.

2.2 Central assumptions

A special feature of process tracing is the importance causal mechanisms in the method. This is in clear contrast to a classical positivist perspective in the theory of science on the one hand, and to many constructivist approaches on the other, both of which - albeit for different reasons - are characterised by a scepticism towards causal statements. Positivism, based on David Hume, only accepts statements about empirical regularities, but does not want to derive any causal statements from them.[4]And constructivist approaches view causal statements with scepticism, especially if they are formulated as social regularities (Yanow 2006; Schwartz-Shea 2006). Furthermore, social causes are not simply seen as part of the external world, but often as an expression of socially characterised meaning. Attributions of causality can thus themselves become the object of investigation.

For some time nowcausality has returned to the political science mainstream, on the one hand via the diversions of economics with its "causal identification revolution" and design-based quantitative research (Morgan 2016) and on the other via process tracing. Although the two approaches have developed independently of each other, they are linked by a criticism of conventional quantitative methods, particularly their inability to distinguish between causal and non-causal effects.

[4]The version of positivism that is still authoritative for parts of the social sciences, the deductive-nomological explanatory model or "covering law" model of Hempel and Oppenheim (1948), follows Hume and systematically avoids statements of causality.

to distinguish between real (spurious) regularities. However, while the causal identification methods are based on experimental research (see Jan-Kowski and Tepe in this volume), process tracing is based more on forms of explanation in historical science and causality concepts of scientific-theoretical realism (Psillos 1999). Here - also in contrast to experimental research - the focus is not on the *effect* of an independent variable on a dependent variable, but on the causal *mechanism* that produces this effect.

But what does mechanism actually mean? Mechanisms link cause and effect, that much is generally recognised. What exactly "connecting" means and what ontological status of the mechanism itself is, however, remains controversial (see Gerring 2007b; Machamer et al. 2000; Glennan and Illari 2018). In any case, mechanisms are not to be equated with "intervening variables"; rather, they themselves generate the effect within a specific setting or context (cf. Falleti and Lynch 2009). One controversial issue is whether mechanisms are ultimately unobservable or whether the black box of causality can be "opened" (authors such as Beach and Pedersen 2012 are particularly optimistic here). However, most proponents of process tracing also assume that the mechanism cannot observed as a whole, but must be inferred indirectly to a large extent. Accordingly, certain observations of states and events can understood as empirical implications of a mechanism (traces), especially if they occur in the context and sequence that is theoretically expected (Nullmeier 2021, pp. 194-196). The connection between the theoretical specification of a mechanism and the observation of implications can be theory-testing or

-The following explains how this can be understood.

2.3 Research objective: Three types of process tracing

With regard to the research objective, a distinction is usually made between three types of process tracing (based on Beach and Pedersen 2012; Bennett and Checkel 2015b): (1) deductive theory-testing analyses, (2) theory- or hypothesis-generating analyses and (3) explanations of specific outcomes. The respective research objective in turn determines the logic of case selection (see below) and the timing of the research process.

In the first type of process tracing, in contrast to classic hypothesis testing in quantitative research, hypotheses are not primarily tested with regard to expected effects, but rather a large number of implications of a hypothesis to be tested - the "links" of the causal chain - are formulated by deduction, which forms the basis of the empirical analysis (see Panke 2012). In addition to proving the expected characteristics of independent and dependent variables, as many of these implications of the mechanism as possible must be found - some authors even demand empirical proof of the complete causal chain (Beach and Pedersen 2012, p. 5). Sometimes it is not so much the explanatory variables disputed in the literature, but rather which mechanism the explanation. A classic example is the

"democratic peace", where it is not the empirical correlation - between democratic dyads and peace - but the mechanism that is disputed (Russett 1994; see George and Bennett 2005, Ch. 2; Müller 2002). In this situation, process tracing is particularly helpful because it allows alternative mechanisms to be tested against each other within one or more cases.

Theory-generating process tracing begins without a plausible explanation for an outcome. The aim is to reconstruct the process between cause(s) and outcome and to inductively capture a possible abstract mechanism and its trigger and relate it to the existing repertoire of theories. Of course, this does not happen "out of the blue", but requires an analytical toolbox of concepts, typologies, models and mechanisms with which the observations be categorised theoretically. Due to the strong influence of Popper's falsificationism in the social sciences, the literature on systematic theory generation is clearly underdeveloped in comparison to theory testing, and systematic introductions are still lacking (but see Swedberg 2011). In both theory-testing and theory-generating process tracing, the conclusions clearly go beyond the cases studied, i.e. they are - at least to a certain extent - generalised.[5]Cases are thus seen as comparable under certain conditions; they are not *sui generis* phenomena; they are understood as "cases of something" (see below).

Although the explanation of specific outcomes, the third possible research objective, is not concerned with generalisation beyond the cases studied, the theoretical mechanisms of explanation are also made explicit here - in contrast to many historiographical analyses (see Elman and Elman 2001). It is therefore not "atheoretical" in the sense of Lijphart (1971, pp. 691-693),[6]but has an "idiographic" flavour in the sense of the distinction between "idiographic" and "nomothetic" research (Windelband 1894), as the individual event takes centre stage (Gerring 2006b). The aim is not to describe the object as completely as possible, but rather to provide a "minimally sufficient explanation" of an event or condition that combines systematic and idiosyncratic explanatory factors and mechanisms (Beach and Pedersen 2012, Chapter 4), whereby discrimination is made between alternative explanations.[7]A case is often initially identified as a "puzzle" from the perspective of research or everyday knowledge, often as "deviant case". One example from policy research is the case-oriented explanation of crises or "policy disasters" (Sa- gan 1995), such as Diane Vaughan's detailed study of the Challenger disaster

[5]Statements about the strength or specific form of an effect cannot usually be made in the context of process analyses.

[6]It is also doubtful whether there can even be such a thing as "atheoretical" research in the face of "theory-laden" data.

[7]Minimally sufficient is an explanation "that accounts for all of the important aspects of an outcome with no redundant parts being present" (Beach and Pedersen 2012, p. 18). However, it cannot be ruled out that in other cases a different explanation applies to the outcome; it is therefore not necessary.

(2009). The aim is to understand how such disasters could occur. Investigation reports in the wake of scandals and similar events often follow this pattern.

As is so often the case, the boundaries between the three types of process tracing are fluid. The more not only purely case-specific, idiosyncratic, but also generally applicable, abstract explanatory factors are used to explain specific outcomes, the more likely it is that the results will take on a theorising or even testing character. In policy research, in contrast to historical science, this is the normal case.

2.4 Case¨ lle and case selection

Deliberate case selection is a prerequisite for studies of fewer cases or even individual case studies to outthemselves in this form. A prerequisite for case selection is the so-called "casing" (Ragin 1992; see also Rohlfing 2012, pp. 24-28), i.e. the answer to the question: What is a case and what is it a case of? The casing brings together considerations of research interest, population, data and theoretical frame of reference. In many situations, casing is relatively unproblematic, as proven research conventions and standardised data exist. Sometimes, however, the conventions are unusable for new questions or there is no agreement on the appropriate unit of investigation and the appropriate analytical level. Explicit "casing" is unavoidable here. For an investigation of socio-political reforms, for example, either entire countries with their welfare state configurations, individual programmes (e.g. pensions, unemployment insurance), individual legislative proposals or policy instruments can be the cases of investigation.

Once it is clear how the cases are to be constituted, next step is decide which specific cases are to included in the study. There are a number of good practices here.[8]The need for (deliberate) case selection arises from the fact that process tracing is based on time-consuming qualitative data collection and analysis methods that rarely allow more than a handful of cases to be included. Instead of random selection, process tracing is based on logical or Bayesian[9]inference, as will be explained in the next section. Only very few situations are conceivable,

[8]This article only presents those methods of case selection that are specifically relevant to process analysis. Methods whose conclusions are on "cross-case analysis" such as Mill's methods, are not included.

[9]Bayesian inference is based on Bayes' theorem for calculating conditional probabilities.
probabilities. Bayes' theorem makes it possible to infer other probabilities (e.g. explanations for these observations) on the basis of known, sometimes subjective probabilities (e.g. from empirical observations) or to carry out an "update" of the latter. Bayesian methods in statistics are usually contrasted with "frequentist" methods, which most of the quantitative methods currently used in political science.

in which no case selection needs to be made at all, for example if it is a kind of "full survey" of an extremely rare phenomenon. However, not all case selection methods are based exclusively on positive cases (i.e. the presence of an outcome of interest on the dependent variable); often negative cases (where the outcome is not present) must also be analysed (Mahoney and Goertz 2004) or it is initially unclear whether the cases are positive or negative, as the selection is based on the independent variable.

2.5 Case selection methods

Case selection methods can be roughly categorised into three groups (see also Seawright and Gerring 2008, following Rohlfing 2012), p. 1) methods based on empirical memory distributions, 2) methods based on empirical correlations and 3) methods based primarily on theoretical expectations (Table 1).

The first category includes methods in which cases are selected that are characterised by "special" values on the dependent variable or one or more independent variables. This includes in particular the *extreme case method*. The extreme case method is theory-generating. For this purpose, a case is selected that has extremely high or low values on an independent *or* dependent variable of interest. The distribution of characteristics within the population must be known, at least approximately (for example, through knowledge of the secondary literature on the topic or quantitative indicators). One chooses
z. For example, a country in which radical reforms took place, on the assumption that equally clear changes also took place in the relevant independent variable and - this is central to process tracing - a mechanism is clearly demonstrable. Alternatively, an extreme value on an independent variable can be selected and its effects analysed (Seawright 2016) or, if the relationship is already known from other studies, an extreme combination of cause and effect can be selected with the aim of identifying the mechanism.

Table 1 Classification of case selection

Basis for case selection	Method	Selection on X and/or Y	Research orientation
Distribution	Extreme case	X or Y	Theory generating
	Various cases	X	Theory-generating or -testing
	Various cases	Y	Theory generating
Context	Typical case	X/Y context	Theory-generating or -testing
	Outlier (deviant case)	X/Y context	Theory generating
Theory	Most-likely	X	Theory testing
	Least-likely	X	Theory testing

observe. A disadvantage of the extreme case method is the possibility of u-shaped or other non-linear correlations, which may render the logic of the extreme case method invalid. Although it should still be possible to observe mechanisms within the case, it is no longer so easy to draw conclusions about other cases: while, assuming a linear relationship, the investigation of an extremely low and extremely high case at least suggests conclusions about the cases in between, this no longer applies in the case of non-linearity. This possibility should therefore always be taken into consideration. Furthermore, it should be considered whether the selected cases are possibly deviating cases or even non-comparable cases that fall outside the relevant population.

The *diverse cases method* (Seawright and Gerring 2008) is based on the idea that several cases that together cover as much of the distribution of the variables as possible tend to be more representative (even if they do not form a random sample). The conflict of objectives between the number of cases and representativeness is therefore directly taken into account here. It is conceivable to select three cases, one with an extremely low value on an independent (or dependent) variable, one with a high value and one average case. For theory-generating process tracing, the selection can also be made on the dependent variable (King et al. 1994, pp. 141-142). A slightly different justification for the *selection of* diverse *cases* is the requirement to prove causal mechanisms in as many different contexts as possible in order to increase the validity of the conclusions (Levi-Faur 2006). In this case, context variables (e.g. sector, region, culture) rather than independent and dependent variables serve as the basis for case selection.

In the case of *typical cases* or their opposite, the *deviating cases* (outliers), the selection is made on the basis of a known empirical correlation. This can, but does not have to, be done as part of a mixed-method design (Rohlfing and Starke 2013; Lieberman 2005). A typical is one that is particularly well explained by existing theories. However, the mechanism or mechanisms are often unclear or alternative mechanisms exist in the literature. The typical case is therefore used to work out mechanisms that can be generalised through process analysis (theory series) - which can then be tested on further cases or new data material - or to discriminate between mechanisms (theory test). Process tracing of deviant cases (see Lijphart 1971) is clearly theory-generating, as both the mechanism and the independent variable are unknown, which ensure that the case is not (well) explained. However, at the beginning of the investigation it is not clear whether the cause of the deviation is purely contingent and case-specific or transferable to other cases and potentially generalisable. In other words, it is often not clear at the outset whether the analysis is ultimately
is "only" an explanation of a specific case or theory-generating.

Finally, types of case selection should be mentioned that are primarily theory-based. In *most-likely cases*, there are strong theoretical reasons to expect a mechanism or a certain outcome. Paul Pierson (1994), for example, the Reagan and Thatcher governments for his study of social policy cuts, as against the background of party difference theory

(as well as the declarations of intent of the actors). Conversely, these expectations are particularly weak in a *least-likely case*.[10]The joke of this method is that its results are of particular interest when expectations are *not* fulfilled, e.g. when existing theories a certain reform, but surprisingly nothing has happened. It remains controversial whether theories can actually be tested effectively with the help of individual *most-likely* or least-likely *cases* (Gerring 2007a; Eckstein 1975). Process tracing should therefore not stop at the statement of the "wrong" outcome, but should also examine the possible alternative explanations and mechanisms behind it. Pierson, for example, takes the absence of radical cuts as the starting point for his approach of the "new politics of the welfare state".

3 Data basis: Systematics beyond standardisation

3.1 Process observations as a data basis

Process tracing of a single case should not be misunderstood as a non-determined N ¼ 1 study. As a within-case analysis, it is always based on a large number of data points and thus benefits, in the words of the methodologist Donald Campbell, from the "discipline coming from a richness of relevant details" (1975, p. 175). In practice, therefore, it is not a lack of data that is problematic, but rather its abundance and disorganised form.

While in quantitative analyses and sometimes also in some case-comparative designs the standardisation of the data must be at the beginning, process tracing allows greater freedom here. At this point, the distinction between "data-set observations" (DSOs) and "causal-process observations" (CPOs) is often used (Collier et al. 2010, pp. 182-196). CPOs are something other than a value or a row of values in a tabular data set (DSOs); rather, they denote "an insight or piece of data that provides information about context or mechanism" (Collier et al. 2010, p. 184). Even if causality itself is not observable, the analysis of processes and thus the temporality of the data remains central (Grzymala-Busse 2011). George and McKe- own (1985) already emphasised that in process tracing it is not the individual data point but "a stream of behaviour through time" (1985, p. 36) that has theoretical value. In other words, it is about having the most reliable information possible about the sequence and order of events, their duration and speed, the actors involved and their relative positions and resources, expectations before and assessments after key events, and (informal or formal) decision-making processes,

[10]Sometimes the "*crucial case*" is also mentioned in this context, but this is merely a particularly strong variant of the most-likely/least-likely *design* (Rohlfing 2012; Eckstein 1975; Gerring 2007a).

relevant and discarded options and the like. These observations do not have to be independent of each other - which is absurd for processes anyway - nor are they equally weighted. On the contrary, in process tracing a few observations are often given special weight for the argumentation (see next section).

The data format is also diverse. A particular strength of process tracing lies in its fundamental openness to completely different types of data in terms of their degree of standardisation, scale level, source, public availability, etc. Secondary and primary data and their collection through field research are generally permissible and desirable. Moreover, the data used is not limited to qualitative formats; rather, it is possible and even advantageous if quantitative data, if available, is also used in primarily qualitative case studies (Maxwell 2010).[11]In part, the approach is similar to triangulation, in which different data formats and sources are also combined in order to increase reliability (see Siewert i. d. B.), but it is less formalised. It is even possible that within a study the same (part of a) mechanism is measured differently in different cases, depending on what is appropriate - i.e. valid - for the respective case.

In order to obtain a sufficient database for process tracing, the network must therefore cast very wide. This approach is usually compared to that of a detective (Collier 2011). A detective tries to match as many available observations of a case as possible with the theoretical implications of alternative explanations - comparable to the suspects in a criminal case - and thus arrive at a coherent conclusion (see also next section). The contrast with highly standardised quantitative (but also some qualitative) procedures could hardly be greater, yet a high degree of systematicity can be achieved, albeit primarily through the dialogue between theory and empiricism, i.e. the close link between thinking about the implications of (alternative) causal mechanisms and empirical observation: "All pieces of evidence are relevant to the central argument (they are not "random"), but they do not comprise observations in a larger sample", according to Gerring (2006a, p. 178). The downside of the detective method is the great effort required to collect data for convincing process tracing (George and McKeown 1985, p. 41) (Pierson 1994).

3.2 Secondary¨ analysis and primary¨ survey

Process analytical research moves between secondary analysis and primary research, between the evaluation of documents and field research. The practice of analysing the existing historical literature for process tracing by secondary analysis

[11]Only when quantitative-statistical evaluation methods are also used is this referred to as a "mixed methods" design.

"however, is widely used (Mahoney and Villegas 2007). The advantage is that several cases can be compared, sometimes over longer periods of time. In terms of the disciplinary division of labour, one relies on the judgement of country or period experts in order to work in a more theoretically oriented manner and avoid the high time and financial costs of conducting one's own research (Skocpol 1984). This can work well, but it doesn't have to. The worst case is the uncritical "cherry picking" of individual historical observations that have no validity outside their original context of analysis (Lustick 1996). It can also happen that (for reason) distorted representations of historians adopted uncritically, thus reproducing their *bias*. To increase the validity and reliability of historical secondary analyses, Cameron Thies suggests a few basic rules, including: if possible, always consult several authors, work chronologically from newer to older publications and the original research context of the supposedly neutrally reported facts (Thies 2002).

Due to the problems mentioned, it is advisable to always at least supplement the secondary analytical data generation with primary analysis. The most important methods are archival research and interviews, often in the context of extensive field research (for practical aspects, see Lieberman et al. 2004). Archival research has a shadowy existence in the literature on political science methods (Frisch et al. 2012 is an exception). This is surprising insofar as many political scientists deal with processes that can be traced in files (if one knows how to find them). In political science education, the basics of archival studies are an exception and so interested students and researchers inevitably have to learn these methods themselves.[12]The increasing availability of documents online is an opportunity, but it harbours the risk of distortion if no additional search is made for material that is exclusively available "offline". It is easy to imagine that certain political actors are highly selective in the (online) publication of documents with a view to their own publicity. Primary sources must therefore be actively sought and edited. However, this should not obscure the fact that the source situation itself is also characterised by historical and political factors:

> Historical evidence is drastically and irremediably incomplete; it is often of dubious validity; and it tends to be biased - favouring the victors rather than the losers, the lasting developments rather than the historical dead ends and detours, the rich and the educated rather than the poor and illiterate, and so forth (Rueschemeyer and Stephens 1997, p. 56; see also Thies 2002).

[12]Dobson and Ziemann, for example, offer an overview of the challenges of different types of texts, such as speeches, letters and autobiographical sources (Dobson and Ziemann 2009). In German, there are a number of introductions to archival work for historians, some of which are also relevant for political scientists (Beck and Henning 2012; Brenner-Wilczek et al. 2006).

3.3 Interviews

Qualitative interviews are much better established as a form of primary research than archival research. In addition to the general literature on qualitative interviews from the neighbouring disciplines of sociology, ethnology and psychology, the specialist literature on expert and elite interviews is particularly relevant for political scientists (Mosley 2013; Dexter 2006 [1970]; Bogner et al. 2002; Leech 2002; Goldberg and Hildebrandt 2020).[13]A number of general difficulties of interviews - power asymmetries, manipulation, *hindsight bias* - can be particularly pronounced in interviews with decision-makers and interest groups. The selection of participants is made deliberately - i.e. not by random selection - and is based on contextual knowledge about decision-making processes (Lynch 2013; Tansey 2007). Longer, semi-structured interviews are primarily suitable for process tracing. There are different approaches to the evaluation. It is often a matter of reconstructing manifest events and processes with the help of participants or contextual knowledge (e.g. on a political field). Here, the evaluation can relatively *ad hoc*. The situation is similar with illustrative quotations. However, if characteristics such as preferences, argumentation methods and positions of several actors are to be collected, systematic coding of the interviews with the help of qualitative analysis software (e.g. MAXQDA, Atlas.ti, NVivo) makes sense. At this point, however, a better reception of interpretative social science - from which such evaluation methods originate - through causally arguing process tracing would be desirable (for a helpful categorisation, see Gläser and Laudel 2013). This form of systematic analysis of interviews and other primary data is still uncommon in process tracing.

4 Data analysis: Trace reading as a method

4.1 Van Evera's four types of empirical evidence

So how exactly does convincing process tracing emerge from the diverse traces that have been gathered by detective work? Are there specific evaluation rules? It is clear that in process tracing, the sheer number of observation points does not say anything about the quality of the conclusions. Rather, individual pieces of evidence can take on particularly great weight. The proximity to criminal investigations and legal evidence, but also to clinical diagnostics, is also evident here. The current debate now centres on what exactly the weighting of

[13]The literature on qualitative interviews in general is very extensive and ranges from theoretical to more practically orientated treatises (Gläser and Laudel 2010; Rubin and Rubin 2011; Rathbun 2008; Harvey 2011).

Tab. 2 4 Types of tests

		Certainty (certainty)	
		High	Low
Uniqueness	High	"Doubly-decisive test"	"Smoking gun test"
	Low	"Hoop test"	"Straw-in-the-wind test"

Source: own illustration based on Van Evera (1997)

[14]Van Evera (1997) made an important initial contribution with his idealised distinction between four types of qualitative tests (see Table 2): "straw-in-the-wind test", "hoop test", "smoking gun test" and "doubly-decisive test". The four tests differ in the degree of "certainty", i.e. the probability with which a hypothesis is correct if certain evidence is available, and in the degree of their "uniqueness"[(15)], i.e. the probability with which the alternative hypothesis is automatically correct in the absence of evidence. The latter depends on whether the empirical predictions are made by several theories simultaneously (Zaks 2017). This typology sounds confusing at first, which has to do in particular with the fact that in some cases the existence and non-existence of evidence can justify differently strong conclusions. I will therefore go into the four different "tests" in more detail below.[16]It is important to note that certainty and selectivity each refer to more than one alternative explanation/hypothesis of an outcome. The two dimensions can also be understood in the terminology of necessary and sufficient condition (Rohlfing 2012, pp. 182-183; Mahoney 2012).

The "*straw-in-the-wind test*" is the weakest of the four tests, as it is characterised neither by certainty nor by selectivity. Van Evera (1997, p. 32) cites the lack of a written Führer order on the Holocaust as an example. Neither is such a document necessary to argue that Hitler personally ordered the Holocaust; nor does its absence automatically allow strong conclusions to be drawn about Hitler's lack of support or about alternative historical causes for the Nazi policy of extermination. If such a document were found, it would certainly have its place in the argument, but would not be sufficient to invalidate alternative explanations.

In the "*hoop test*", observations are certain, but not conclusive. Negative evidence can invalidate a hypothesis, but positive evidence has very little significance. The example here is the presence of the suspect in the city where a crime took place. If a suspect can prove that he

[14]It is controversial whether and how exactly counterfactual inferences should be used in process analyses (Lebow 2010). While George and Bennett are critical of such inferences (2005, pp. 167-170), authors such as Rohlfing argue in favour of them systematically (2012, pp. 175-178).

[15]I would like to thank Henning Deters for suggesting this translation.

[16]To add to confusion, the four "tests" can be used not only in theory-testing analyses are used. The term "test" should be understood here more as a diagnostic test or test procedure.

was out of the country, he is probably innocent, but his presence itself only proves that it was possible for him to the offence, not that he it.

The situation is exactly the opposite with a "*smoking gun test*". The low certainty and high selectivity of such a test mean that while positive evidence can be decisive in favouring a hypothesis over its alternatives, a negative finding is not. The suspect with the smoking gun was probably the perpetrator. However, just because a suspect is not holding a smoking gun does not mean that he can be considered exonerated.

The *"doubly-decisive*" test is comparable to footage of a crime (e.g. a bank robbery) from a surveillance camera. On the one hand, positive evidence (e.g. a clearly recognisable face) is highly incriminating (¼ high certainty), on the other hand it can considered exculpatory if the accused cannot be seen (¼ selectivity).

Van Evera's four test forms have subsequently been discussed in detail.[17]Bennett's Bayesian-inspired interpretation is particularly illuminating (Bennett 2008b, 2015). Analogous to Bayesian reasoning with conditional probabilities, inference in process tracing is seen as an "updating" of a priori probabilities (see also Beach and Pedersen 2012). Put simply, the more surprising or unique observations are from the perspective of certain hypotheses, general case knowledge and the theoretical expectations of the literature, the greater their weight in the argumentation. In contrast to falsificationist notions of tests, Bayesian tests assume that by observing certain evidence we can "update" the (alternative) hypotheses with which we approach a case. This updating often depends on how established a theory is and how "surprising" (to put it simply) the empirical evidence found is against the background of the existing literature. A previously very well established theory (high prior probability) will only marginally invalidated by new contradictory evidence, but how surprising and novel this evidence is plays a major role.[18]The aim of process tracing is now to find suitable evidence for the strongest possible tests, e.g. failed "*hoop tests*", positive "*smoking gun tests*" and perhaps even a "*doubly decisive test*".[19]To this end, it is essential that the empirical implications of alternative explanations (including their mechanisms!) systematically worked out and compared with each other in order to capture their theoretical uniqueness.[20]What are the empirical implications of alternative explanations?

[17]However, the confusion as to how exactly the four tests are to be understood persists in the literature and the interpretations of Bennett, Rohlfing, Mahoney, Beach and Pedersen and Collier differ in detail.

[18]It becomes somewhat more complicated when updating in direct relation to alternative explanations.

[19]This does not mean that "*straw-in-the-wind tests*" are worthless. However, the lack of strength of the test must be compensated for by the number of empirical observations.

[20]Campell already recommends that authors of case studies record the theoretically expected implications in detail in addition to field notes (1975, p. 186).

Expectations if theory A is correct? But also: What should *not* be observed? How surprising are certain observations? Are there individual observations that can be used to clearly discriminate between theory A and B? The more systematically and transparently such questions are dealt with, the better. In their textbook, Beach and Pedersen (2012) describe the various forms of Bayesian inference in detail and explain with many examples how to recognise particularly strong inferences and how to use them. In recent years, a highly technical debate has developed within the Bayesian-oriented process tracing literature, in which, among other things, attempts have been made to assign quantified probabilities to qualitative inferences again (Fairfield and Charman 2017). It is doubtful whether this is possible at all and whether the effort involved is in proportion to the yield (Zaks 2021). In any case, "trace reading" in the context of process tracing can now be carried out in a differentiated and transparent manner if one has sufficiently specific theories and detailed empirical knowledge.

4.2 The presentation of process-analytical findings

So far, little attention has been paid to the presentation of the results of process tracing. Even in policy research, case studies are usually presented in narrative form. In contrast to historical approaches and the stereotype of the "storytelling", however, there is great resistance to narrative research among some methodologists (Beach and Pedersen 2012; Tilly 1999). However, if one understands a narrative, as is common in literary studies, as a form of *representation of* a story (a "plot"), this resistance is not understandable. This is because the plot can be highly abstracted and theory-led (a good example is the analytic narratives of Bates et al. 1998; Levi 2002). Due to the importance of causal mechanisms, narration is even very compatible with process tracing, because both are about selectively representing processes and events that exhibit temporality and causality and take place against the background of a context (the "setting" of the story) (see also Abbott 2008). Therefore, there is hardly any way around narrative representations. However, there are few indications of what constitutes a "well-told" case study (but see Büthe 2002; Ruback 2010).

5 Applications in policy research

5.1 Theory of process tracing: Steinmo (1989)

The first example of (implicit) process tracing in policy research is Sven Steinmo's study of tax policy in the UK, Sweden and the USA (Steinmo 1989). Steinmo begins by distinguishing three approaches to explaining country differences in the level and structure of taxation: "'interests,' 'values'

and 'the state'" (1989, p. 500). The first approach refers to theories that focus on parties and interest organisations, the second on cultural explanations and political culture and the third on (historical) institutionalist theories. In the following, we will test which of the three approaches can best explain the differences between the countries. The starting point is the counter-intuitive empirical observation that the USA has traditionally favoured redistributive taxes and corporate taxes more than Sweden (Great Britain lies between these two poles). Added to this is the extraordinary degree of fragmentation in the US tax system. Steinmo's analysis is based on secondary sources, revenue statistics and a series of interviews with tax policy actors. His empirical analysis describes the general patterns of tax policy in the three countries, sometimes in a rather stylised form: pluralist politics in the US, corporatist in Sweden and party politics in the UK. These patterns are mainly attributed to historical differences in political institutions (especially electoral systems and structures of interest organisations). Individual tax reforms, on the other hand, are used more for illustrative purposes. Steinmo compares the preferences of the actors, the processes of tax legislation and the structural outcomes with the expectations of the theoretical approaches. Mechanisms are not formulated abstractly and deductively, but rather *ad hoc*. However, the author is aware of the methodological limitations of his study and writes

> A more complete explanation would require a more fully elaborated analysis of the development of American political institutions over time and the dynamic interaction of these institutions and the large, fragmented, and expanding economy. Unfortunately, space does not allow us to delve into this discussion here (Steinmo 1989, p. 511).

However, Steinmo's study is not based exclusively on three unrelated case studies, but is rather a kind of mixed-methods design with elements of both process tracing and a small-N comparison. Not only the developments *within* the cases, but also the contrasts *between* the cases therefore play a role in the complex overall argumentation. His theory test shows that the institutionalist approach can explain the structural differences and typical decision-making patterns in tax policy better than interest- or value-orientated approaches alone.

5.2 Theory-generating process tracing: Trampusch (2006)

Another example of process tracing is Christine Trampusch's study of the social policy reform policies of the Schröder government (Trampusch 2006). She contrasts a traditionally incrementalist view, which focuses on the many possible blockades in the political system of the Federal Republic and sees (party-political) preferences as largely static and exogenous, with an approach that (also) recognises preferences as endogenous through process tracing.

The first approach - which she sees as being influenced by the feedback effects of the policies. The first approach - from her
"interaction analysis" - has difficulties explaining the red-green social reforms; they appear to be a deviating case both from the usual incremental policy pattern of the Federal Republic and generally from social reforms in systems with many veto points. The author now shows empirically that the reform process did not correspond to the assumptions of the interaction model in central points, but rather to a historical-institutionalist model of endogenous preference change. Exogenous shocks, for example, which interaction analyses usually use to explain radical policy change, played a certain role, but are not sufficient to explain the change. Instead, Trampusch develops a complex narrative in which the interaction perspective is supplemented by a sequence-orientated perspective in which feedback and preference changes play a central role.

Methodologically, Trampusch's study corresponds to process tracing in many respects. The focus is not only on correlations between independent and dependent variables across several cases, but also on processes within a case and the sequence of central, theoretically significant events. It uses the status of German social policy under the red-green coalition as a deviant case to contribute to the further development of existing theories of policy change. It explicitly aims to modify, rather than falsify, interaction theory. In addition to secondary analyses, primary sources (e.g. coalition agreements, statements) and media reports are also used to substantiate preferences and decision-making processes. However, this analysis also deviates from the ideal model of process tracing when it comes to specifying causal mechanisms. Instead of linking alternative causes and outcomes with clearly formulated theoretical mechanisms and continuous causal chains, Trampusch uses the two theoretical perspectives more as heuristics (2006, p. 63) and examines the extent to which empirical policy processes are more similar to one or the other rather general policy model in order to finally arrive at a synthesis of the approaches.

5.3 Explanation¨ of a specific outcome: Mos (2014)

A good example of historically oriented process tracing to explain a specific outcome is Martijn Mos' analysis of the anti-discrimination clause regarding sexual orientation in Article 13 the 1997 Amsterdam Treaty (Mos 2014). This event was surprising in that, at the time of the treaty negotiations, only a small minority of Member States had implemented comparable clauses at national level and there was no great interest at intergovernmental level in extending non-discrimination protection to other groups. Mos shows that the European Parliament was able to play an important role here, which contradicts the dominant liberal-intergovernmentalist and institutionalist declarations of the Amsterdam Treaty. It traces in detail how members of the Parliament have been able to play a role since the

In the early 1980s, gay rights *activists* brought gay rights onto the supranational agenda (and kept them there) despite the lack of a basis of competence and were ultimately able to influence the treaty negotiations together with *gay rights activists*. He draws on official documents, his own archival research and 33 semi-structured interviews with representatives of EU institutions, member states and interest groups.

Mos himself describes his analysis as theory-generating process tracing (2014, p. 634). In contrast to Trampusch's theoretical claim, however, Mos' explanation appears to be driven far more by the surprising empirical case than by the theoretical puzzle. He emphasises that his case study

> does not call for a wholesale rejection of leading theoretical perspectives, but [… .] suggests that overall grand theory should coexist with an appreciation of the more contextual dynamics of specific episodes of treatymaking (Mos 2014, p. 634).

The result of the analysis is not a new theory, but rather two conditions under which the European Parliament could become a gateway for anti-discrimination rights: "agenda overload" and specific "issue characteristics". Whether these conditions are transferable to other policies remains to be seen. As is so often the case, the difference between the forms of process tracing in practice is difficult to determine.

What all three examples have in common is that the explanatory mechanisms are not formulated in a textbook manner together with all the links in the causal chain. In this respect, they are similar to the "pattern matching" method, which also examines whether individual, sometimes relatively loosely connected empirical aspects fit different theories or not. Even if not all parts of the explanation can ultimately be proven empirically, a more trans- parent theoretical specification would help. A typical weakness of process tracing analyses is the unequal treatment of alternative explanations. This is also reflected in the - otherwise exemplary - examples. For example, the "culturalist" perspective is treated rather neglected by Steinmo. However, this is a missed opportunity to fully utilise one of the potential strengths of process tracing.

6 Strengths¨ strengths, weaknesses¨ and gji te criteria of process tracing

One of the strengths of process tracing lies in the possibility of reaching causal conclusions about alternative explanations, even in a small number of cases, by empirically analysing causal mechanisms. Secondly, the high degree of flexibility with regard to the data sources and formats that are included must be mentioned as a strength. Thirdly, the methods for drawing conclusions on the basis of sometimes few but significant observations are now quite sophisticated.

One critical point to note is that theories in policy research are often not specified to the extent that would be necessary for a strong deduction of empirical implications. In particular, the theoretical mechanisms are often only vaguely outlined, are controversial or even completely absent. Arguments that certain factors - such as certain actors - "play a role" are not uncommon. Theoretical tests on detailed mechanisms, which are usually the focus of the literature on process tracing, would therefore often be virtually impossible.

This point can be formulated even more fundamentally. Although process tracing is placed close to scientific realism, certain aspects are still characterised by positivism. In particular, the most theory-orientated variant is based on the hypothetico-deductivist model of scientific knowledge, according to which theories contain a series of relatively unproblematic deducible implications that can be translated into basic empirical observations and thus tested. The problem, however, is that this derivation and translation is itself theory-driven (Schindler 2013). Ideally, before a theory test can be carried out, these "auxiliary theories" would have to be tested themselves (and in turn their auxiliary theories beforehand). More pragmatic alternatives to the hypothetical-deductivist model are presented by Clarke and Primo (2012) for political science (see also Friedrichs and Kra- tochwil 2009). Another alternative is offered by "abductive" approaches in social science (Reichertz 2013). Their ideal of reading traces and iteratively refining explanations based on individual or a few cases comes very close to the practice of process tracing without wanting to "test" (see Fairfield and Charman 2019). After all, Bennett and Checkel also emphasise that process tracing is usually a mixture of deduction and induction.

Finally, an unresolved problem is the question of the generalisation of process-analytical findings. Quantitative research can refer to classical probability theory in its attempt to generalise from a sample to the population. Process tracing (still) lacks this clear foundation. On the continuum between pure individual observation and research into universal laws, process tracing claims the area of "contingent generalisation" (George and Bennett 2005; see also Mitchell 2002). George and Bennett understand this as a type of generalisation that includes contextual conditions and generalises only within a comparable group of cases (e.g. developed welfare states, ethnically mixed societies, resource-rich autocracies). The comparability of the context can be derived from similar historical developments, cultural or geographical circumstances. However, it is unclear exactly what this means. What are the criteria that tell us whether and to what extent our conclusions can also claim external validity? An alternative concept is "generalisation by mechanism" (Bengtsson and Hert- ting 2014; Ylikoski 2019), which assumes that, under certain conditions, findings about causal mechanisms can be generalised from one context to another.

Regardless of these open questions of scientific theory, not all process tracing is the same and the debate about appropriate methodological quality criteria is in full swing. Bennett and Checkel, for example, set out 10 criteria for good process tracing (see box, see also the checklist in the appendix by Beach and Pedersen 2012), which emphasise the aspect of alternative explanations, among other things. This makes sense, because a risk to the validity of process tracing arises from psychological effects such as confirmation bias or specific problems of Bayesian inference, such as the "*base rate fallacy*" (Bar-Hillel 1980), which leads to the "certainty" and "discriminatory power" of evidence being misjudged. Here it can help to always include alternative explanations systematically and on an equal footing in empirical analyses.

Box 1: Guidance criteria for process tracing

1) Cast the net widely for alternative explanations
2) Be equally tough on the alternative explanations
3) Consider the potential biases of evidentiary sources
4) Take into account whether the case is most or least likely for alternative explanations
5) Make a justifiable decision on when to start
6) Be relentless in gathering diverse and relevant evidence, but make a justifiable decision on when to stop
7) Combine process tracing with case comparisons when useful for the research goal and feasible
8) Be open to inductive insights
9) Use deduction to ask "if my explanation is true, what will be the specific process leading to the outcome?"
10) Remember that conclusive process tracing is good, but not all good process tracing is conclusive

Source: Bennett and Checkel (2015b)

7 Conclusion

Process tracing is characterised by a specific approach to theories, data and conclusions, which was presented in this article. The method is suitable for many questions of policy research. Characteristics such as case orientation, causal explanation, process character, actor-centredness, theory pluralism and data pluralism are good reasons for the traditionally strong affinity between policy research and process tracing. At the same time, policy research is methodologically a pluralistic sub-discipline. Therefore, the differentiation of the methodological discussion should not lead to the points of contact between the methodological approaches being lost sight of (see, for example, Nullmeier 2019 on interpretative methods). A stronger dialogue with quantitative research, QCA users and interpretative-qualitative research, not only on scientific theoretical foundations, but also on actual research practice, is necessary.

more important than ever. This is because when it comes to the validity problems of interview data or case selection problems, for example, the gaps between the different camps are often far less deep than expected. There is also the expectation that mixed methods designs can combine the strengths of the different methods in a complementary way. Here too, process tracing can play an important role by emphasising causal mechanisms and detailed case knowledge.

Literature

Abbott, H. Porter. 2008. *The Cambridge introduction to narrative*. Cambridge: Cambridge University Press.

Bar-Hillel, Maya. 1980. The base-rate fallacy in probability judgements. *Acta Psychologica* 44(3): 211-233.

Bates, Robert, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast, eds. 1998. *analytic narratives*. Princeton: Princeton University Press.

Beach, Derek, and Rasmus B. Pedersen. 2012. *process-tracing methods: Foundations and guidelines*. Ann Arbor: University of Michigan Press.

Beach, Derek, and Rasmus Brun Pedersen. 2016. *causal case study methods: Foundations and guidelines for comparing, matching, and tracing*. Ann Arbor: University of Michigan Press.

Beck, Friedrich, and Eckart Henning, eds. 2012 *The archival sources. With an introduction to the auxiliary historical sciences*, 5th ed. Vienna/Cologne/Weimar: Böhlau.

Bengtsson, Bo, and Nils Hertting. 2014 Generalisation by mechanism: Thin rationality and ideal-type analysis in case study research. *Philosophy of the social sciences* 44(6): 707-732.

Bennett, Andrew. 2008a. Building communities, bridging gaps: Alexander George's contributions to research methods. *Political Psychology* 29(4): 489-507.

Bennett, Andrew. 2008b. Process tracing: A Bayesian perspective. In *The Oxford handbook of political methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier, 702-721. Oxford: Oxford University Press.

Bennett, Andrew. 2015. Appendix: Disciplining our conjectures. Systematising process tracing with Bayesian analysis. In *Process tracing: From metaphor to analytic tool*, eds Andrew Bennett and Jeffrey T. Checkel, 276-298. Cambridge: Cambridge University Press.

Bennett, Andrew, and Jeffrey T. Checkel, eds. 2015a. *Process tracing. From metaphor to analytic tool*. Cambridge: Cambridge University Press.

Bennett, Andrew, and Jeffrey T. Checkel. 2015b. Process tracing: From philosophical roots to best practices. In *Process tracing. From metaphor to analytic tool*, eds Andrew Bennett and Jeffrey T. Checkel, 3-37. Cambridge: Cambridge University Press.

Bogner, Alexander, Beate Littig, and Wolfgang Menz, eds. 2002: *The expert interview: Theo- ry, method, application*. Opladen/Wiesbaden: Leske + Budrich/Springer.

Brenner-Wilczek, Sabine, Gertrude Cepl-Kaufmann, and Max Plassmann. 2006. *introduction modern archival work*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Büthe, Tim. 2002. Taking temporality seriously: Modelling history and the use of narratives as evidence. *American Political Science Review* 96(3): 481-493.

Campbell, Donald T. 1975. "Degrees of freedom" and the case study. *Comparative Political Studies* 8(2): 178-193.

Checkel, Jeffrey T. 2008. process tracing. In *Qualitative methods in international relations. A pluralist guide*, ed. Audie Klotz and Deepa Prakash, 114-127. Heidelberg/Berlin: Springer.

Clarke, Kevin A., and David M. Primo. 2012. *A model discipline: Political science and the logic of representations*. Oxford: Oxford University Press.

Collier, David. 2011. Understanding process tracing. *PS: Political Science & Politics* 44(4): 823-830.

Collier, David, Henry E. Brady, and Jason Seawright. 2010. sources of leverage in causal inference: Toward an alternative view of methodology. In *Rethinking social inquiry*, ed. Henry E. Brady and David Collier, 2nd ed. 161-199. Lanham: Rowman & Littlefield.

Dexter, Lewis A. 2006 [1970]. *Elite and specialised interviewing, ECPR classics*. Colchester: ECPR Press.

Dobson, Miriam, and Benjamin Ziemann, eds. 2009. *Reading primary sources: The interpretation of texts from nineteenth- and twentieth-century history*. London: Routledge.

Eckstein, Harry. 1975. case study and theory in political science. In *Handbook of political science*, ed. Fred I. Greenstein and Nelson W. Polsby, 79-137. Reading: Addison-Wesley.

Einhorn, Hillel J., Don N. Kleinmuntz, and Benjamin Kleinmuntz. 1979. linear regression and process-tracing models of judgement. *Psychological Review* 86(5): 465.

Elman, C., and M. F. Elman, eds. 2001. *Bridges and boundaries: Historians, political scientists, and the study of international relations*. Cambridge, MA: MIT Press.

Fairfield, Tasha, and Andrew Charman. 2019. A dialogue with the data: The Bayesian foundations of iterative research in qualitative social science. *Perspectives on Politics* 17(1): 154-167.

Fairfield, Tasha, and Andrew E. Charman. 2017. Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis* 25(3): 363-380.

Falleti, Tulia G., and Julia F. Lynch. 2009. context and causal mechanisms in political analysis. *Comparative Political Studies* 42(9): 1143-1166.

Ford, J. Kevin, Neal Schmitt, Susan L. Schechtman, Brian M. Hults, and Mary L. Doherty. 1989. process tracing methods - Contributions, problems, and neglected research questions. *Organizational Behaviour and Human Decision Processes* 43(1): 75-117.

Friedrichs, Jörg, and Friedrich Kratochwil. 2009. On acting and knowing: How pragmatism can advance international relations research and methodology. *International Organisation* 63(4): 701-731.

Frisch, Scott A., Douglas B. Harris, Sean Q. Kelly, and David C. W. Parker. 2012. *Doing archival research in political science*. New York: Cambria Press.

George, Alexander L. 1979. Case studies and theory development: The method of structured, focused comparison. In *Diplomacy: New approaches in history, theory and policy*, ed. G. Lauren, 43-68. New York: The Free Press.

George, Alexander L., and Andrew Bennett. 2005. *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.

George, Alexander L., and Timothy J. McKeown. 1985. case studies and theories of organizational decision-making. *Advances in Information Processing in Organisations* 2:21-58.

Gerring, John. 2006a. *Case study research: Principles and practices*. Cambridge: Cambridge University Press.

Gerring, John. 2006b. Single-outcome studies: A methodological primer. *International Sociology* 21(5): 707-734.

Gerring, John. 2007a. Is there a (viable) crucial-case method? *Comparative Political Studies* 40(3): 231-253.

Gerring, John. 2007b. The mechanismic worldview: Thinking inside the box. *British Journal of Political Science* 38(1): 161-179.

Gläser, Jochen, and Grit Laudel. 2010. *expert interviews and qualitative content analysis*, 4th ed. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gläser, Jochen, and Grit Laudel. 2013. Life with and without coding: Two methods for early-stage data analysis in qualitative research aiming at causal explanations. *Forum Qualitative Sozial- forschung/Forum Qualitative Social Research* 14(2). https://doi.org/10.17169/fqs-14.2.1886.

Glennan, Stuart, and Phyllis Illari, eds. 2018. *The Routledge handbook of mechanisms and mechanical philosophy*. London: Routledge.

Goldberg, Felix, and Achim Hildebrandt. 2020. expert interviews. In *Handbuch Methoden der Politikwissenschaft*, ed. Claudius Wagemann, Achim Goerres and Markus B. Siewert, 267-284. Berlin/Heidelberg: Springer.

Grzymala-Busse, Anna. 2011. time will tell? Temporality and the analysis of causal mechanisms and processes. *Comparative Political Studies* 44(9): 1267-1297.

Hall, Peter A. 2006. Systematic process analysis: When and how to use it. *European Management Review* 3(1): 24-31.

Harvey, William S. 2011. Strategies for conducting elite interviews. *Qualitative Research* 11(4): 431-441.

Hempel, Carl Gustav, and Paul Oppenheim. 1948. studies in the logic of explanation. *Philosophy of Science* 15(2): 135-175.

Hogarth, Robin M. 1974 Process tracing in clinical judgement. *Behavioural Science* 19(5): 298-313.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton: Princeton University Press.

Kühberger, Anton, Michael Schulte-Mecklenbeck, and Rob Ranyard. 2011. introduction: Windows for understanding the mind. in *A handbook of process tracing methods for decision research: A critical review and user's guide*, ed. Michael Schulte-Mecklenbeck, Anton Kühberger, and Rob Ranyard, 1-18. New York: Psychology Press.

Lebow, Richard N. 2010. *Forbidden fruit: Counterfactuals and international relations*. Princeton: Princeton University Press.

Leech, B. L. 2002. Asking questions: Techniques for semistructured interviews. *PS: Political Science & Politics* 35(4): 665-668.

Levi, Margaret. 2002. modelling complex historical processes with analytic narratives. In *Actors - Mechanisms - Models. Zur Theoriefähigkeit makro-sozialer Analysen*, ed. Renate Mayntz, 108-127. Frankfurt a. M./New York: Campus.

Levi-Faur, David. 2006. Varieties of regulatory capitalism: Getting the most out of the comparative method. *Governance* 19(3): 367-382.

Lieberman, Evan S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99(3): 435-452.

Lieberman, Evan S., Marc Morje Howard, and Julia Lynch. 2004. symposium: Field research. *Qualitative Methods: Newsletter of the American Political Science Association Organised Section on Qualitative Methods* 2(1): 2-15.

Lieberson, Stanley. 1991. Small N's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases. *Social Forces* 70(2): 307-320.

Lijphart, Arend. 1971. comparative politics and the comparative method. *American Political Science Review* 65(3): 682-693.

Lustick, Ian S. 1996. History, historiography, and political science: Multiple historical records and the problem of selction bias. *American Political Science Review* 90(3): 605-618.

Lynch, Julia F. 2013. Aligning sampling strategies with analytic goals. In *Interview research in political science*, ed. Layna Mosley, 31-44. Ithaca: Cornell University Press.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. thinking about mechanisms. *Philosophy of Science* 67(1): 1-25.

Mahoney, James. 2003. strategies of causal assessment in comparative historical analysis. In *Comparative historical analysis in the social sciences*, ed. James Mahoney and Dietrich Rueschemeyer, 337-372. Cambridge: Cambridge University Press.

Mahoney, James. 2012. the logic of process tracing tests in the social sciences. *Sociological Methods & Research* 41(4): 570-597.

Mahoney, James, and Gary Goertz. 2004. the possibility principle: Choosing negative cases in comparative research. *American Political Science Review* 98(4): 653-669.

Mahoney, James, and Claudio M. Villegas. 2007. Historical enquiry and comparative politics. In *The Oxford handbook of comparative politics*, ed. Carles Boix and Susan Stokes, 73-89. Oxford: Oxford University Press.

Maxwell, Joseph A. 2010. Using numbers in qualitative research. *Qualitative Inquiry* 16(6): 475-482.

Mayntz, Renate. 2002. On the theoretical capacity of macro-social analyses. In *Actors - Mechanisms - Models. Zur Theoriefähigkeit makro-sozialer Analysen*, ed. Renate Mayntz, 7-43. Frankfurt/ New York: Campus.

Mill, John Stuart. 1968 [1872]. *System der deduktiven und induktiven Logik: Eine Darlegung der Grundsätze der Beweislehre und der Methoden wissenschaftlicher Forschung, Gesammelte Werke, transl. and ed. by Theodor Gomperz, vol. 1*. 2nd, German ed., repr. of the Leipzig edition, 1884 ed. Aalen: Scientia.

Mitchell, Sandra D. 2002 Contingent generalisations: Lessons from biology. In *Akteure - Mechanismen - Modelle: Zur Theoriefähigkeit makro-sozialer Analysen*, ed. Mayntz, 179-195. Frankfurt a. M./New York: Campus.

Morgan, Kimberly J. 2016. Process tracing and the causal identification revolution. *New Political Economy* 21(5): 489-492.

Mos, Martijn. 2014 Of gay rights and Christmas ornaments: The political history of sexual orientation non-discrimination in the treaty of Amsterdam. *JCMS: Journal of Common Market Studies* 52(3): 632-649.

Mosley, Layna, ed. 2013. *Interview research in political science*. Ithaca: Cornell University Press.

Müller, Harald. 2002. antinomies of democratic peace. *Political Quarterly* 43(1): 46-81.

Norman, Ludvig. 2015. interpretive process tracing and causal explanations. *APSA Qualitative and Multi-method Analysis Newsletter* 13(2): 4-9.

Nullmeier, Frank. 2019. interpretative policy research and causal mechanisms. *Journal of Political Science* 29(2): 153-171.

Nullmeier, Frank. 2021. *Causal Mechanisms and Process Tracing: Perspectives on Qualitative Policy Research*. Frankfurt/New York: Campus.

Panke, Diana. 2012. process tracing: testing multiple hypotheses with a small number of cases. In *Research design in European studies: Establishing causality in Europeanisation*, ed. Theofanis Exadaktylos and Claudio M. Radaelli, 125-140. Basingstoke: Palgrave Macmillan.

Pierson, Paul. 1994. *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrench- ment*. Cambridge, MA: Cambridge University Press.

Przeworski, Adam, and Henry Teune. 1970 *The logic of comparative social inquiry*. New York: Wiley-Interscience.

Psillos, Stathis. 1999. *scientific realism: How science tracks truth*. London: Routledge.

Ragin, Charles C. 1992. "Casing" and the process of social inquiry. In *What is a case? Exploring the foundations of social inquiry*, ed. Charles C. Ragin and Howard S. Becker, 217-226. Cambridge: Cambridge University Press.

Rathbun, Brian C. 2008. Interviewing and qualitative field methods: Pragmatism and practicalities. In *The Oxford handbook of political methodology*, ed. Janet M. Box-Steffensmeier, Henry
E. Brady and David Collier, 685-701, Oxford: Oxford University Press.

Reichertz, Jo. 2013. *Abduction in qualitative social research. On the discovery of the new*, 2nd ed. Wiesbaden: Springer VS.

Ricks, Jacob I., and Amy H. Liu. 2018. process-tracing research designs: A practical guide. *PS: Political Science & Politics* 51(4): 842-846.

Rohlfing, Ingo. 2012. *case studies and causal inference: An integrative framework*. Houndmills/ Basingstoke: Palgrave Macmillan.

Rohlfing, Ingo, and Peter Starke. 2013. building on solid ground: Robust case selection in multi-method research. *Swiss Political Science Review* 19(4): 492-512.

Ruback, Timothy J. 2010. 'Let me tell the story straight on:' *Middlemarch*, Process-tracing methods, and the politics of narrative. *British Journal of Politics and International Relations* 12(4): 477-497.

Rubin, Herbert J., and Irene S. Rubin. 2011. *qualitative interviewing: The art of hearing data*. London/New York: Sage.

Rueschemeyer, Dietrich, and John D. Stephens. 1997. comparing historical sequences - A powerful tool for causal analysis. *Comparative Social Research* 16(1): 55-72.
Russett, Bruce. 1994 *Grasping the democratic peace: Principles for a post-Cold War world*. Princeton: Princeton University Press.
Sagan, Scott. 1995. *the limits of safety: organisations, accidents, and nuclear weapons*. Princeton: Princeton University Press.
Schindler, Samuel. 2013. Observation and theory-ladenness. In *Encyclopedia of philosophy and the social sciences*, ed. Byron Kaldis, 695-698. London: Sage.
Schwartz-Shea, Peregrine. 2006. judging quality: Evaluative criteria and epistemic communities. In *Interpretation and method: Empirical research methods and the interpretive turn*, ed. Dwora Yanow and Peregrine Schwartz-Shea, 89-113. Armonk/London: M.E. Sharpe.
Seawright, Jason. 2016. The case for selecting cases that are deviant or extreme on the independent variable. *Sociological Methods & Research* 45(3): 493-525.
Seawright, Jason, and John Gerring. 2008. case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly* 61(2): 294-308.
Skocpol, Theda. 1979. *States and social revolutions: A comparative analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
Skocpol, Theda. 1984. Emerging agendas and recurrent strategies in historical sociology. In *Vision and method in historical sociology*, ed. Theda Skocpol, 356-391. Cambridge: Cambidge University Press.
Steinmo, Sven. 1989. Political institutions and tax policy in the United States, Sweden, and Britain. *World Politics* 41(4): 500-535.
Swedberg, Richard. 2011. theorising in sociology and social science: Turning to the context of discovery. *Theory and Society* 41(1): 1-40.
Tansey, Oisín. 2007. process tracing and elite interviewing: A case for non-probability sampling. *PS: Political Science & Politics* 40(4): 765-772.
Thies, Cameron. 2002. A pragmatic guide to qualitative historical analysis in the study of interna- tional relations. *International Studies Perspectives* 3(3): 351-372.
Tilly, Charles. 1999. The trouble with stories. In *The social worlds of higher education. Handbook for teaching in a new century*, ed. Ronald Aminzade and Bernice Pescosolido, 256-270. Thousand Oaks: Pine Forge Press.
Trampusch, Christine. 2006. sequence-orientated policy analysis. *Berlin Journal of Sociology* 16(1): 55-76.
Trampusch, Christine, and Bruno Palier. 2016. Between X and Y: How process tracing contributes to opening the black box of causality. *New political economy* 21(5): 437-454.
Van Evera, Stephen. 1997. *guide to methods for students of political science*. Ithaca: Cornell University Press.
Vaughan, Diane. 2009. *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.
Windelband, Wilhelm. 1894. history and natural science. Rede zum Antritt des Rektorats der Kaiser-Wilhelms-Universität Straßburg, gehalten am 01.05.1894. Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Philosophisch-Historische Klasse; Jg. 1910, Abh. 14. https://doi.org/10.11588/diglit.20767. Accessed on 06.04.2022.
Yanow, Dvora. 2006. introduction. In *Interpretation and method: Empirical research methods and the interpretive turn*, ed. Dwora Yanow and Peregrine Schwartz-Shea, xi-xxvi. Armonk/ London: M.E. Sharpe.
Ylikoski, Petri. 2019. Mechanism-based theorizing and generalisation from case studies. *Studies in History and Philosophy of Science Part A* 78:14-22.
Zaks, Sherry. 2017. relationships among rivals (RAR) a framework for analysing contending hypotheses in process tracing. *Political Analysis* 25(3): 344-362.
Zaks, Sherry. 2021 Updating Bayesian(s): A critical evaluation of Bayesian process tracing. *Political Analysis* 29(1): 58-74.