



AARHUS UNIVERSITY



Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Wauters, B. and Beach, D. (2018). Process tracing and congruence analysis to support theory-based impact evaluation. *Evaluation*, 24(3), 284-305. DOI: 10.1177/1356389018786081

Publication metadata

Title:	Process tracing and congruence analysis to support theory-based impact evaluation
Author(s):	Benedict Wauters and Derek Beach
Journal:	<i>Evaluation</i> , 24(3), 284-305
DOI/Link:	https://doi.org/10.1177/1356389018786081
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Process tracing and congruence analysis to support theory based impact evaluation.

Abstract

In this article, a theory based impact evaluation of a European Social Fund intervention is developed further to demonstrate the difference between process tracing and congruence analysis and their relative (dis)advantages.

Keywords

Process tracing, congruence analysis, theory based impact evaluation, theory of change

Authors:

Benedict Wauters

Flemish Department of work and social economy, Belgium

Derek Beach

Department of political science, Aarhus University, Denmark

Corresponding author:

Benedict Wauters, Tommeltlaan 32, 2640 Mortselsel, Belgium.

E-mail: benedict.wauters@wse.vlaanderen.be

Tel. 32 476940208

Benedict Wauters is in charge of evaluation, innovation and methodology at the European Social Fund department in Flemish Department of work and social economy, Belgium. He is also a lecturer in research methods and public sector innovation at Antwerp Management School as well as an independent expert.

Introduction

Theory based impact evaluation has been defined (EC, 2013, page 51-52) as “an approach in which attention is paid to theories of policy makers, programme managers or other stakeholders, i.e., collections of assumptions, and hypotheses - empirically testable - that are logically linked together...”.¹

The article responds to the need for “additional exemplars of theory-driven evaluations, including reports of successes and failures, methods and analytic techniques, and evaluation outcomes and consequences” (Coryn et al 2011). Hence, an example of an evaluation conducted in the European Social Fund in Flanders (De Rick et al, 2014), will be used to illustrate the differences between process tracing (as described in Beach and Pedersen, 2013 and 2016) and congruence analysis (as described in Blatter and Haverland, 2012), including relative advantages and disadvantages. Both process tracing and congruence analysis are analytical strategies developed in social science that can be used for causal inference within a single case study and that can be adapted to work in theory based impact evaluation, as proposed by Stern et al (2012). However, little guidance exists as to what strategy to choose and how to execute this in an evaluation context. Hence, a comparison of what they entail and insight into relative (dis)advantages should be of interest to practitioners.

A theory based impact evaluation case from Flanders as a basis for comparing different approaches

The Personal Development Process initiative

In the context of the evaluation of the European Social Fund (ESF) programme in Flanders, an impact evaluation was conducted of one specific intervention within this programme. The intervention is referred to as a “Personal Development Process” (PDP) as described in De Rick et al (2014). The study defines this PDP as a supportive process with as its goal to improve the labor market oriented personal development of individuals.

The PDP was open to unemployed as well as employed Flemish citizens, but we will focus in this article only on employed persons. For the latter, the PDP essentially entails that a coach works with a participant to help them define how they would like their career to evolve in the future and to take appropriate action to make this a reality. The reason for supporting this with public finances is that it is assumed that the participant will become more pro-active in shaping their career, which will make them more self-reliant in the face of misfortunes such as sudden massive lay-offs.

The initial approach that was followed for this evaluation was based on Chen (2005, 2006). This approach was chosen as the aim was to develop more insight into how the intervention worked (if at all). In line with this approach, a distinction was made between an action model and a change model. The action model is meant to represent a systematic plan for arranging staff, resources, setting and support organizations to reach target populations and provide intervention services. The change model visualizes how the implementation of the intervention will affect determinants, which in turn, will change the outcomes.

¹ An overview of the history of theory based evaluation is provided by Coryn et al (2011) and does not have to be repeated here.

The change model hence corresponds to an “outcome chain” and the action model to the “action theory” as described by Funnel and Rogers (2011). An outcomes chain is to be distinguished from a full “theory of change” as the latter also integrates the action theory as well as non-intervention factors such as broad context factors (socio-economic, political...), government policies, rules, activities of partners and other actors, public opinion groups and media, intervention critics, characteristics of target groups,... that are assumed to be present for the outcomes that the intervention aims to achieve to be able to materialize (p. 222).

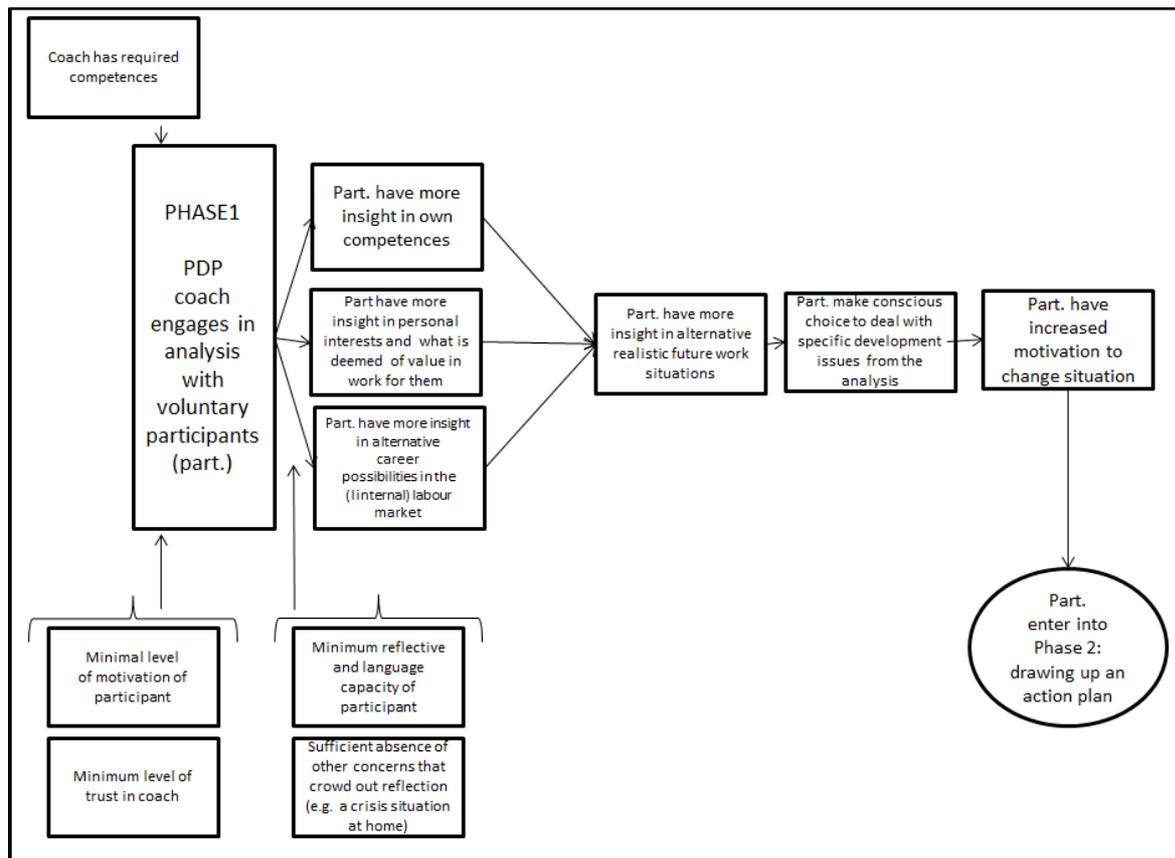
For the PDP evaluation, expectations were drawn up in the action model regarding the following key dimensions:

- philosophy of the PDP
- PDP cycle
- PDP documentation
- the nature of the PDP support
- requirements for the PDP coach
- characteristics of the target group
- overall organization of PDPs

The change model, based on the same preliminary, exploratory research as the action model, is depicted below in two figures to improve its readability.

Figure 1 shows what is assumed to occur based on a first (phase 1) coaching intervention with the aim to identify which competences participants should develop to realize their ambitions in terms of their career. They do this by first gaining insight into competences they already possess, what they find interesting in work (for which they do not necessarily possess competences yet) and what possibilities exists for them in their environment. These insights combine into a realistic expectation that they should be motivated to pursue.

Figure 1: detail of phase 1 of the PDP cycle

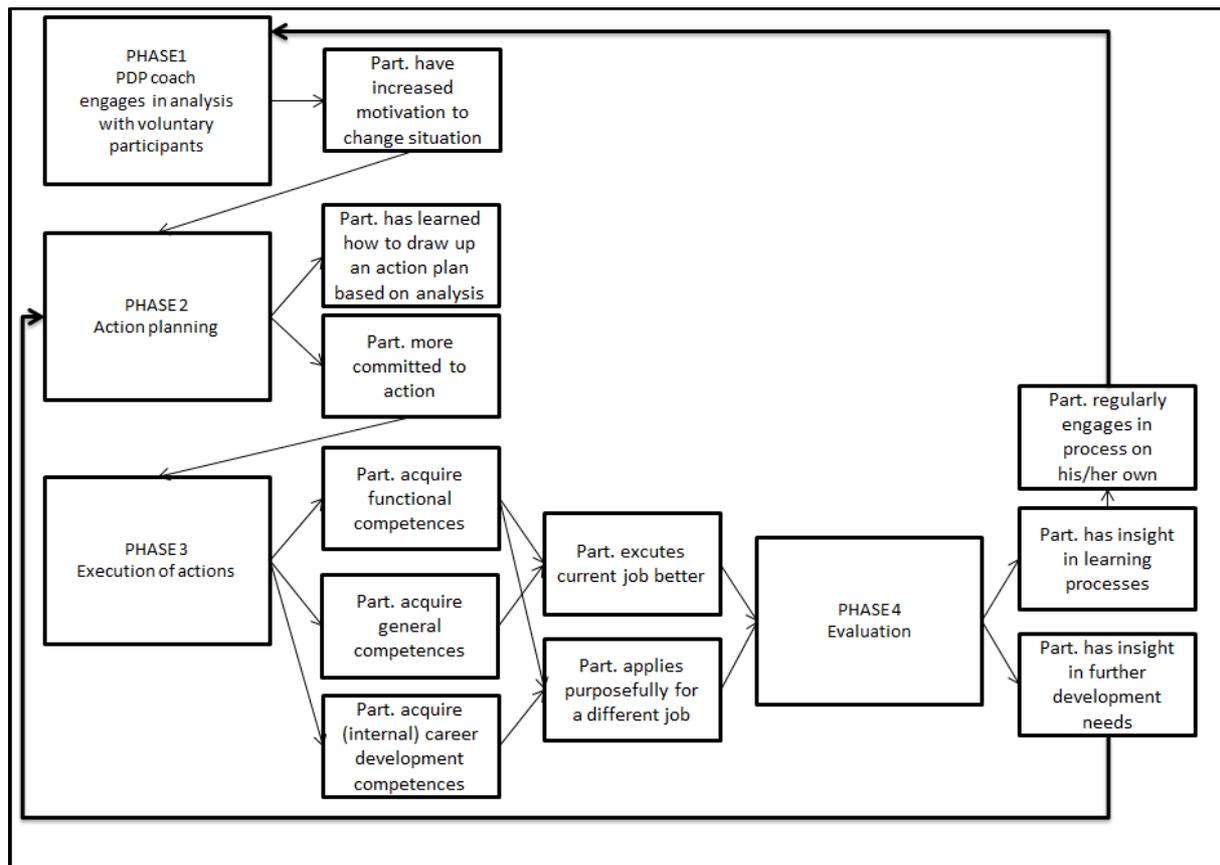


Adapted from De Rick et al (2014)

Figure 1 also depicts several conditions outside the intervention e.g. the assumed absence of other concerns that could crowd out reflection during the PDP. These conditions therefore correspond to non-intervention factors as described by Funnel and Rogers (2011) earlier. Hence, the action model and the change model together with these conditions can be said to describe a theory of change.

Figure 2 shows that after having identified their development issues, participants are expected to move on towards action planning, execution of these actions and evaluation of this execution. It also shows that once the outcomes of phase 4 are achieved, a virtuous circle of regularly going back into phase 1 and 2, autonomously, is expected to occur, indicating that the participant has become more pro-active in shaping their career.

Figure 2: overall PDP cycle.



Adapted from De Rick et al (2014)

It should be noted that the action model (depicted by the four “phases” in figure 2) interacts with the change model in four different places. This means that some changes are presupposed to have occurred before parts of the action model are to become relevant (e.g. without motivation to change the situation, derived from participation in phase 1, there would be no entering phase 2).

Conducting the evaluation

The evaluation set out to assess to what extent the action model had been respected and to what extent this created the hoped-for change as depicted in the change model. In this article, the focus (for illustrative purposes) will be on only one case, namely of the PDP process as organized within and by a private sector company for its own employees. Within this case, data collection consisted of semi-structured interviews with the coordinator of the POP project, with four participants and with three coaches of these participants. These interviews were transcribed and analyzed with NVIVO qualitative data analysis software. The evaluators were, next to Chen (2005, 2006), also instructed to use the analytical approaches proposed by Miles and Huberman (2014)².

Findings of the evaluation

In terms of assessing the fidelity of this case to the action model, the evaluators found that it deviated in various aspects from what was theoretically intended . For example, in the case there

² In this sense, the article should not be read as a criticism of their work as they were not instructed to use process tracing or congruence analysis.

was no attention to exploring job possibilities beyond the current or already decided future position and there was very little exploration of a participants' motivation and qualities apart from those relevant to the predefined current or future job.

Another important deviation from the theoretical concept was that a demand-led approach, where the participant is central, was not followed. Also, phase 1 and 2 (analysis and action plan) were taken together in only one conversation. Furthermore, the execution of the action plan was not supported with more conversations. The evaluation phase (12 months later) was rather part of a new cycle starting, instead of being a self-standing action. Essentially, in this case, it became clear that the PDP concept was narrowed to a traditional HR planning cycle where training is identified based on shortfalls in terms of competence profile applicable to a specific position. The company perspective was dominant in all of this.

In terms of the change model, we will focus for illustrative purposes only on the effects of executing phase 1, expected to occur before moving into phase 2. In any case, the evaluators state that, given the deviations from the intended action model, phase one of the PDP should not be expected to create anything else for the participants but insights in their own competences. They conclude that indeed, moderate (due to the restricted scope) levels of insight into competences are obtained .

The next sections will use this evaluation to illustrate in the first instance process tracing and afterwards congruence analysis.

What if we had used process tracing for this study?

According to Beach and Pedersen (2013, p. 2 and p. 28) process tracing is a within-case study method for making causal claims based on a mechanistic and deterministic view of causality. In relation to evaluation, process tracing aims to explain how an intervention has worked in real world cases. The mechanistic element of process tracing implies that a "causal mechanism" needs to be theorized as a process –described as an unbroken chain of action and reaction (activities) enacted by entities (actors) – that connects the potential cause with its hypothesized outcome. As such, it concerns a "process whereby causal forces are transmitted through a series of interlocking parts of a mechanism to produce an outcome" (Beach and Pedersen, 2013, p. 40). Interlocking can be easily misunderstood as merely connected in some way. However, we would put forward that the difference is quite crucial: the former assumes that we make an unbroken chain of activity explicit, the latter does not. From this it is also clear that the intervention X is not the mechanism. It only sets the mechanism into motion as a process. This corresponds to authors such as Pawson (2003, p. 473) who state that "pathway from resource to reasoning is referred to as the programme mechanism".

Schmitt and Beach (2015) develop such a chain for a part of the reasoning behind budget support in the context of development aid. Another example, concerning Gender Responsive Budgeting, is provided by Bamanyaki and Holvoet(2016).

Similar to Pawson and Tilley (1997, p. 58), mechanisms are typically understood as being sensitive to the context in which they operate, which means that the same intervention in a different context can be expected to produce different results.

Mechanisms are not theorized to be necessary or sufficient for an outcome; they are 'merely' links between a cause X, that can be necessary or sufficient, and an outcome Y. At the same time, a

specified mechanism is not necessarily the only route from a given X to Y. Various mechanisms can create the connection (separately or in combination), regardless of whether X is sufficient or necessary for Y. In any case, Beach and Pedersen (2016, p. 176) state that it is not possible to use within case evidence to assess if a cause is sufficient or necessary for an outcome. Cross-case analysis in the form of comparisons is required for this. What process tracing can establish is if a hypothesized mechanism indeed connects X with Y.

Beach and Pedersen (2016) also stress that an understanding of a mechanism as a system implies that “parts have no independent existence...in relation to producing an outcome” (p. 35). However, as put forward by Beach and Pedersen (2016, p. 39-40) this does not imply that if we could remove a part (e.g. as in a counterfactual experiment), the mechanism would not function anymore as another, unknown, alternative part could then become active due to potential redundancy of parts. Therefore, to assess whether a mechanism functioned as theorized, we would not engage in counterfactual-based comparison using a control group to assess the effect of a part (or overall process). Instead, in process tracing we attempt to trace empirically the within-case fingerprints left by the activities of entities for each part of a process. If we find that a part was either not present or functioned differently than we expected, we would not necessarily disconfirm the whole mechanism. For example, if in a study of a two-part mechanism we find that part 1 worked as functioned, but part 2 was not present, but instead part 2b kicked in, then we would have revised our theory. However, if we do not find evidence of any further link after part 1 despite repeated attempts, we would conclude that there is no mechanism that links the cause and outcome together because it breaks down after part 1.

There are several variants of process tracing according to Beach and Pedersen (2013 and 2016). “Theory-testing” involves testing whether a hypothesized mechanism links an intervention with an outcome, whereas “theory-building” is a bottom-up process aimed at finding what mechanism, if any, links an intervention with an outcome. “Outcome explaining” aims to continue evidencing the presence of mechanisms until one is satisfied that together these explain the bulk of the observed outcome in a specific case. Here, there is usually no interest in generalizing across cases.

In impact evaluation, the aim is to demonstrate the presence or absence of a mechanism. Hence both theory-testing as well as outcome explaining process tracing can be of interest. Theory building process tracing can be conducted beforehand to ensure a relevant theoretical mechanism (or set of mechanisms) has been elaborated as a causal process, ready for testing. This could be part of an “evaluability” study to be conducted before the actual impact evaluation. In many cases, such a prior study will indeed be required.

The discussion below will limit itself to the theory testing variation as we focus only on one mechanism that we are, in principle, interested to substantiate across several cases, even though for this article, we will limit ourselves to a discussion regarding only a single case. Process tracing as proposed by Beach and Pedersen (2013, p. 56-60, chapter 6, 7 and 8) then consists of the following steps:

- 1) Elaborating the hypothesized mechanisms;

2) Selecting an appropriate case where the intervention is present, the scope conditions for the mechanism are present, and the outcome is present (typical case) or could at least have been present (deviant case);

3) Substantiating the presence of the mechanism by proposing observable implications for all the steps in the mechanism and gathering the corresponding data.

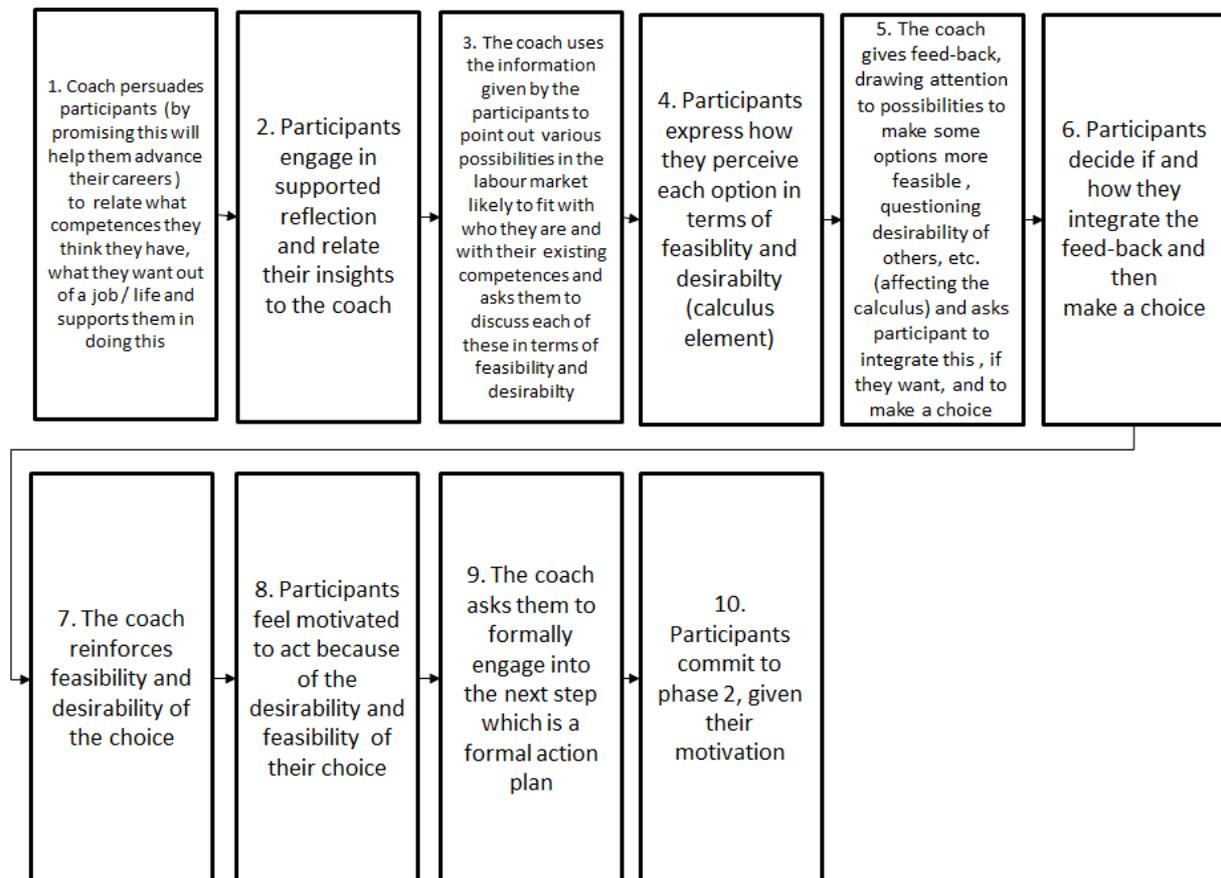
These steps are discussed below.

Elaborating the hypothesized mechanisms in process tracing

As stated above, the first step is to theorise a mechanism as an unbroken chain of action and reaction between various reaction between various actors.

Figure 3 shows what the PDP theory of change (comprising the action and change model) for phase 1 could look like as a mechanism.

Figure 3: rational choice mechanism display of the PDP theory of change



Source: author's own adaptation of De Rick et al (2014)

In keeping with Leeuw (2012), a link is made to existing broader theories, in this case to rational choice theory, as described in Hedström (2005, p.60-66), to make clear the “causal principle” (Cartwright and Hardie, 2012) that is theorized to bind parts of the mechanism together. Clearly, the reaction of participants in the second step (to give information to the coach) depends on them

believing that this reaction brings them closer to something they value (career advancement). They do this by first gaining insight into competences they already possess, what they find interesting in work (for which they do not necessarily possess competences yet) and what possibilities exist for them in their environment. These insights combine into a realistic expectation that they should be motivated to pursue.

Figure 1 where insights lead to choices being made. However, the mechanism display requires to make exactly clear how these insights lead to this choice being made, within the overall rational choice assumption, because we do not want to just assume causal links but instead try to assess them empirically through the activities of agents.

It should also be noted that in a theory-centric approach, only one mechanism at a time is elaborated and tested. It could well be that other mechanisms are implied, e.g. an empowerment mechanism which could be based on Ryan and Deci's (2000) self-determination theory to complement or reinforce the rational choice mechanism. In the theory centric approach, such a mechanism then is also to be fully elaborated and separately studied. In a case-centric, outcome explaining approach, the two mechanisms could rather be studied as one "composite" mechanism.

Selecting cases in process tracing

Beach and Pedersen (2016b) state that, in a first instance, only cases that have achieved a certain threshold of the cause X as well as the outcome Y, along with the expected causally relevant scope conditions, should be selected if the interest is to trace the process from X to outcome Y. As stated earlier, this is referred to as a typical case. Also, cases that are deviant in the sense that they reach the threshold for X and the scope conditions but not for Y can be relevant to detect hidden scope conditions or disabling other causes, but only if there are typical cases to compare with.

In the case of interventions, X is (a part of) an intervention. Hence, it follows that for the PDP we should select cases where X (the PDP action model or parts of it) has been reasonably faithfully implemented as well as achieved the desired (intermediate) outcome Y (for example in the PDP, going ahead with phase 2). The intervention must have reached some threshold value where it is still reasonable to assume that the mechanism can be set in motion. Falling below this value represents critical implementation failure which means we cannot expect that the mechanism will be present hence we should not study the case, regardless of the value of Y. This of course make sense, as indeed the absence of the cause that is supposed to put the mechanism into motion does not bode well for confirming the presence of the mechanism. But also, the absence of any decent Y value also does not bode well for the mechanism. However, a case where X is present and the expected scope conditions were also present, but where Y is not present could be a candidate for theoretical revision aimed at finding unknown causal and/or scope conditions that must be present for the process to work properly.

A first important step in process tracing must therefore be to identify the population of potential cases and to find out whether X and Y and the scope conditions are (sufficiently) present or not. Let us assume that the original phase 1 action model (trigger X) meets the threshold of sufficient implementation, that scope conditions are also met and that there are indeed many participants that decide to go to phase 2 (which be the outcome Y of phase 1). Hence the rational choice mechanism as described in They do this by first gaining insight into competences they already possess, what they find interesting in work (for which they do not necessarily possess competences yet) and what possibilities exists for them in their environment. These insights combine into a realistic expectation that they should be motivated to pursue.

Figure 1 can reasonably be hypothesized to apply.

It should be understood that defining the population of cases is mechanism specific. If we had wanted to follow up on a mechanism founded in self-determination theory, it is possible that a case considered as typical or deviant under rational choice, is labelled differently under self-determination. When researching composite mechanisms in outcome explaining process tracing, we would need to make sure that all the required scope conditions and triggers are indeed present.

Substantiating the presence of the mechanism in process tracing

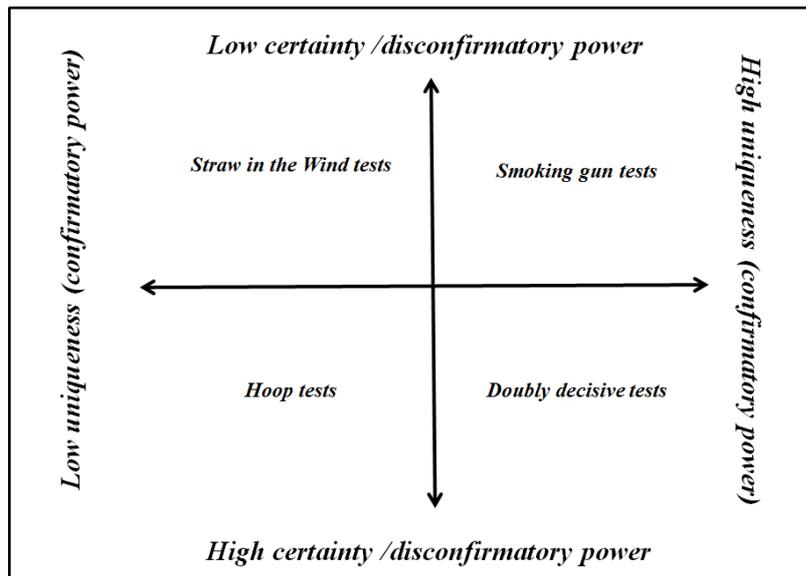
Beach and Pedersen (2013, chapter 6) make clear that we need to formulate empirical expectations regarding the theoretical mechanism as a process: what would we expect to see empirically in a specific case for every single step of the theoretical mechanism conceptualized as an unbroken chain of action and reaction? After that, we need to define and find the actual data that would help us verify if this empirical expectation is met. Such empirical expectations are referred to as “observable implications”.

To ensure an in-depth discussion, let us limit our focus to the claim that participants engage in supported reflection where they give the coach the requested information (the second step in figure 3). For example, in a specific case, we could expect to see *“a written report of the engagement with a coach that states that the participant related insights in their competences to the coach”*. The data to be gathered in this case would obviously be some document. However, we would need to know something about the case already for this observable implication to be useful as a test. Indeed, as having written reports may not have been the norm across all cases, it may be that for such a case a better observable implication would be to expect to hear *“an oral report about insights being gained by participants with the help of the coach”*. The data would then presumably have to be gathered via interviews with coaches and/or participants.

However, the above examples then beg the question what kind of test of the presence/absence of the theoretical mechanism is actually proposed? Beach and Pedersen (2013) provide a powerful way to proceed, depicted in figure 4, based on a framework by Van Evera (1997), which is underpinned by Bayesian probability calculus³.

Figure 4: typology of empirical tests

³ For detailed explanations about Bayesian probability calculus we refer to Beach and Pedersen (2016, p.169 - 214; and Bennett, 2014, p. 266-298).



Adapted from Beach and Pedersen (2013)

The framework is based on two dimensions:

- **Uniqueness:** to what extent can the empirical prediction overlap with a prediction that could be made from other (unspecified) theories? In other words, we are looking for predictions that are very unlikely to hold unless the theory we are investigating is operating. This is also referred to as the confirmatory power of such predictions. Relatively unique predictions can be thought of as ‘signatures’ of the working of a particular part of a process;
- **Certainty:** do we have to find the predicted evidence, which relates to what extent should the prediction be confirmed or we should discard the theory? This is also referred to as the disconfirmatory power of such predictions.

Combining these two dimensions contains four possibilities in terms of the nature of a test and consequently provides a foundation for devising a set of alternative tests. Note that tests can be more or less unique/certain, and in real-world research one test is typically not very informative but needs to be combined with other pieces of independent evidence.

Let us take the oral report on having gained insights given above as an example. If we do not find these kinds of reports, then this would be a serious problem for the proposed theory. Certainty is relatively high here: we really should find this evidence if the theory is to be taken serious, hence not finding it damages the theory.

However, if we do find it, it does not really tell us much because uniqueness is low. Insights could have been derived in many ways that have nothing to do with the support provided by the coach. These insights could have existed already and therefore could be readily presented to the coach without actual reflection. This combination means this is a “hoop” test. We must jump through the hoop or the theory is disconfirmed. But jumping through it does not mean the theory is valid. In other words, merely putting forward that insights were indeed derived from the process with the coach would not enable confirmation of the mechanism. It would have enabled disconfirmation if we had not found it. However, as stated by Beach and Pedersen (2016, p. 287), passing multiple hoop tests can support a theory. While each individual test only has a low level of uniqueness, if they are

independent of other tests, there is an additive effect, meaning that passing many can have a confirmatory effect.

We could make this hoop test much stronger by “tightening” the hoop if we reformulate the observable implication as follows: *“participants relate in an interview exactly how the use of particular tools during the process created new insights -that they did not have before- in terms of their competences”*. Here the uniqueness is higher than before. If we find this evidence, it still lends some confirmatory weight for the theory as it is not so likely this evidence would be equally explained by other theories.

A double decisive test would be delivered by *“direct observation of a participants working with a coach where using tools leads to generating insights that are clearly new for the participant as evidenced by some kind of “aha” experience during the engagement”*. If we find this data, it supports the theory (confirmation), if we do not it falsifies the theory (disconfirmation).

A straw in the wind test would be provided by *“generic documents that mention existing competences for participants as well as the gap with expected competences”*. Uniqueness is low as there could be many reasons (alternative theories) for such documents to exist. Certainty is also low in cases where such documents were not a mandatory part of the PDP exercise. However, in a case where these documents are mandatory, it would weaken the theory if they are not found, making it a hoop test.

Also, if the document is not generic but very specific with a lot of references about the PDP process and how insights were derived from specific exercises, in a case where such documents are not mandatory, then it would be a smoking gun: finding it has confirmatory power (as it is unlikely such content in a document would exist unless for the PDP) but low disconfirming power (not finding it is not a problem as these documents are not necessarily to be found for the theory to remain valid).

Next, as stated by Beach and Pedersen (2013, p. 127-9), the accuracy of the data is also very important. It does not matter how strong a test is, if the data we are gathering to conduct the test regarding an observable implication is not accurate, then the test is useless. If an observable implication in the PDP research was that a participant should be able to relate the insights they derived from a PDP regarding their competences and if these interviews are taking place a very long time after the engagement, recall could be a major issue and this means that if participants cannot relate their insights anymore, this may not demonstrate anything due to inaccuracy. On the other hand, if the interviews are conducted immediately after the coaching conversation and then the participants cannot relate any insights, there is no reason to think accuracy is an issue. In addition, as put forward by Beach and Pedersen (2016, p.190-191), if the evidence is found, one needs to make sure it is representative of the empirical record and not just 'cherry-picked'. If the evidence is not found, one needs to ascertain if this is problematic or not (e.g. it could be that it was just impossible to access it). These are assessments of empirical uniqueness and certainty. Finally, it is needed to aggregate the variety of tests and their results to draw conclusions concerning the theory at hand (p.207-212).

What would have been the (dis)advantage of using process tracing relatively to the original methodology?

Schmitt and Beach (2015) note two main advantages in their reflection on the use of process tracing for evaluation:

- More explicit theorization of causal mechanisms in an intervention logic, giving us a better working knowledge of how processes play out in real-world cases;
- A clearer logic of inference (Bayesian tests), enabling stronger and more transparent inferences to be made.

However, they also note that, given constraints on time and resources, typically an evaluation needs to be more limited in scope, focusing on the connection between only a few elements within a broader intervention logic. This reflects a position taken by Pawson et al (2015), using as an example a policy of public disclosure of health care performance information, to show that there are several theories that aim to explain different elements of how such a policy may work and that one cannot study all of them: “decide upon which combinations and which subset of theories are going to feature on the short list....comprehensive reviews are impossible and...the task is to prioritize and agree on which programme theories are to be inspected” (p. 27-8).

Other authors have pointed out similar (dis)advantages. Punton and Welle (2015) point out that process tracing provides a firm basis for shedding light on why and how an intervention led to change, ex post without needing a control group. However, the approach does not provide an estimate of how important a cause was relative to other causes. In addition, process tracing requires that one can ascertain the outcome of a case and can therefore be conducted properly only at a moment in time when this information has become available. However, using intermediate outcomes may alleviate this practical issue. Many mechanisms will relate to only parts of a case anyway and hence de facto require intermediate outcomes. Finally, they also point out the issue of needing sufficient time and resources to execute the approach.

Befani, D’Errico, Booker and Giuliani (2016) point out mainly advantages relating to the Bayesian logic of inference: assessing and even measuring confidence in causal claims, with a very high level of transparency regarding why pieces of evidence are deemed of value, avoiding evaluator bias (against or for a hypothesis). In addition, they also point to the advantage of being to execute the approach ex post, without needing to interfere in programme design as well as to the close dialogue between theory and evidence.

When reflecting on the PDP evaluation, we can confirm these potential advantages and disadvantages, relatively to the initial approach.

First, in line with Schmitt and Beach’s (2015) first advantage, a key distinction with the approach proposed by Chen (2005, 2006) is that the mechanism as a process relates much more clearly exactly how X leads to Y. This is not merely a matter of detail in terms of steps, but of ensuring the unbrokenness of the chain of action and reaction as a foundation for causal inference, embodying a hypothesized mechanism (for example rational choice). Hence, “Phase 1” is not put separate as containing an action model that creates a chain of outcomes but actions and outcomes are fully integrated in each step of the process. This is also not a mere description of actors acting sequentially. For example, if we state that a first step is that “A coach has a session with a participant” (actor acting), followed by “the participant has more insights” (actor reacting) (as depicted in figure 1) there is still no clear theoretical mechanism because we have not described

what is linking the two together. Indeed, why would a session lead to insights? We have not “explained” this. A participant could just as well sit there and dream about a nice vacation. There must be a “reason” embedded in the mechanism. The reason, from a rational choice perspective, becomes clearer by stating that “a coach persuades participants (by promising this will help them advance their careers) to relate what competences they think they have, what they want out of a job / life and supports them in doing this.” In other words, participants’ belief that their career will be advanced is how they become engaged in the process of supported reflection.

Beach and Pedersen (2013, p. 38) hence rightly put forward that a theory as depicted in They do this by first gaining insight into competences they already possess, what they find interesting in work (for which they do not necessarily possess competences yet) and what possibilities exists for them in their environment. These insights combine into a realistic expectation that they should be motivated to pursue.

Figure 1 “gray-boxes” the mechanism. It is still better than a “black-box” which would only state that engaging in phase one of a PDP leads to engaging into phase 2 but nonetheless covers over exactly how cause leads to effect. Indeed, with each transition, one should ask the question: why would this reaction by an actor follow from a previous action by an actor? This situates process tracing firmly in the range of approaches as proposed in sociology by Hedström (2005), who breaks down mechanisms at the level of individuals into desires, beliefs and opportunities, and in evaluation by Astbury and Leeuw (2010), who speaks of unpacking black boxes.

Without such an unbroken chain of action and reaction, it is not clear what the significance is of finding data that seems to support the various elements in

Figure 1. Even if we find data that insights are gained, it is not possible to guarantee this is connected to the actions of phase 1. There are gaps between phase 1 and the outcomes. However, as put forward by Beach and Pedersen 2016, p. 85-89), there is no reason why mechanisms should always be specified at a particular level (e.g. micro, as done by Hedström, 2005). If a macro-level theorized process (e.g. of groups acting collectively) has empirical manifestations that can be used to trace a process, there is no methodological reason to lower the level of analysis to the micro-level.

Next, in line with Schmitt and Beach's (2015) second advantage, it is clear that using the Bayesian empirical test empirical test framework, a greater number of tests have been devised than are present in the original evaluation. In original evaluation. In addition, it would be more transparent what the value of these tests is. Indeed, if we look at the if we look at the actual research conducted by the original evaluation, the conclusion that insights were indeed gained were indeed gained (an outcome in

Figure 1) is derived from interview data. Examples of evidence used by the evaluators to assert that phase 1 created more insights in existing competences or just confirmed these insights, are cited in De Rick et al (2014, p. 107). For example, the question in a semi-structured interview with a PDP participant “Did you, because of the PDP, for example get more conscious whether this could be a development point for me?” was answered by statements such as:

- “I think you will also look more consciously at your job. But do I see...? Yes, there are always points to develop. There is always something you can work on.”
- “There are a number of things. But those we knew already before. They just came up again.”

From the discussion above, it should be clear that this is a hoop test with a very large hoop. Finding this data does not tell us anything: insights could have been derived in many ways that have nothing to do with the support provided by the coach. Indeed, the exact formulation of the responses seems to point to the fact that the insights were already there before the PDP. Using the Bayesian framework, the evaluators could have tightened the hoop to gain some confirmatory power. In addition, they could have devised other tests, as proposed above. But in any case, they would never have used such data to confirm the effectiveness of the PDP.

As to the disadvantages, or rather, the heavy demands, of process tracing, we can add:

- Process tracing not only requires knowledge of the outcome in the population of cases where the timing of the evaluation may be too soon (a lesser problem as one can always focus on an intermediate outcome), but it also requires knowledge of the extent to which (part of) the intervention achieved a critical threshold of fidelity in terms of execution of a pre-specified model;
- In addition, once suitable cases have been identified, quite a lot of familiarity with the various cases is assumed in general to specify case specific empirical tests;
- Finally, the approach is very sensitive to misspecifications of the mechanism into the theoretical process of action and reaction. This means that we can expect to have to engage in an iterative process of theory building for fine-tuning every step of the theorized process, followed by more testing of these steps. Hence, the scope of research tends to be rather narrow, as we have done with the PDP by limiting ourselves to phase one and to one mechanism (rational choice) only.

These elements translate indeed into significant time and resource requirements as noted by the aforementioned authors.

What if we had used congruence analysis for this study?

Congruence analysis is, according to Beach and Pedersen (2016: p. 269-301, p. 28), based on engaging in within-case analysis without unpacking each part of the causal mechanism as a continuous process linking X and Y together. Instead, theories in congruence often have the form of a “modus operandi” explanation, with key parts described but without a full-fledged theorized causal process (i.e. mechanism) (Beach and Pedersen, 2016, p. 287).

This means that in the absence of evidence of an unbroken chain of cause and effect between X and Y, there is no way of making a strong causal inference about X being linked to Y. No matter how many times we observe a deterministic regularity of, for example, X thus Y, hence pointing towards X

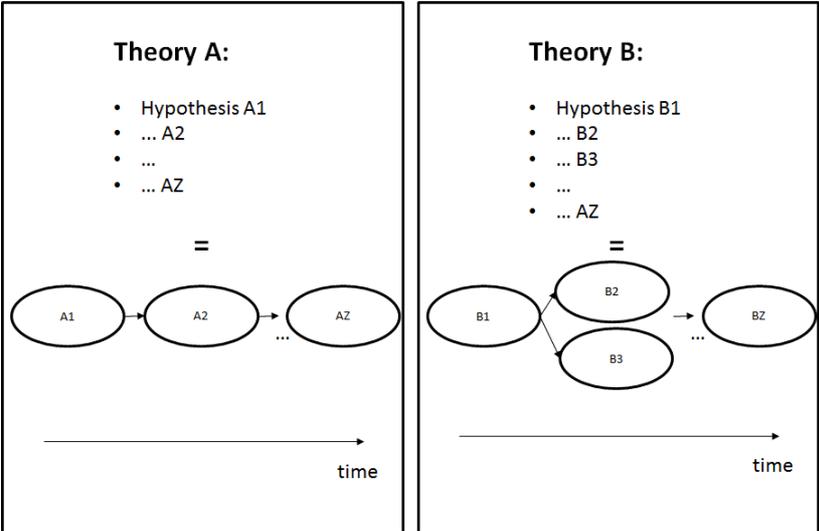
as a sufficient cause, we can never rule out that both Y and X appear because of an alternative cause Z that is sufficient for both Y and X, without there being any direct causal link between Y and X. This is the same issue that has given rise to counterfactual impact evaluation methods (see e.g. Khandker et al, 2010) although these do not conceptualize causality deterministically but probabilistically (in terms of tendencies rather than sufficiency or necessity).

However, what remains possible in the absence of a counterfactual that can be tested, is to compare the strength of the evidence for different theoretical explanations of the outcome Y. It should be emphasized that congruence analysis is not able to definitively “rule out” or “confirm” alternative theories, but rather to establish comparative strength of the evidence for one or more theories. Hence, while Blatter and Haverland (2012, p. 161-2), who have been pioneering this approach, confirm that it is possible to do a single theory congruence analysis, but they also state that this is a weak alternative to using multiple theories.

Different variations of congruence analysis are proposed by Blatter and Haverland (2012, p. 145). Mirroring theory centric process tracing, there could be an interest in researching if an explanation travels across cases. Then we could be interested in understanding which of several theories would be better supported by evidence in explaining a range of cases (referred to as the competing theory approach). However, similar to outcome explaining process tracing, we might be interested in exploring how causes work together to produce an outcome in one particular case (referred to as complementing theory approach), stopping when the evidence becomes too weak and when there are no obvious theories left, with no intention to generalize.

It should be noted that Blatter and Haverland (2012, p. 163-4) do not require to depict theory as is done in figure 5. However, the visualization of when propositions occur over time can clearly add value because the sequence in which they occur can then become a relevant observable that can be tested empirically in a case. However, this is not like process tracing where a mechanism is depicted as an unbroken chain of action and reaction. In fact, figure 5 is closer to a traditional depiction of a theory of change as in figure 1 and 2.

Figure 5: a theory as composed of a rich set of loosely linked hypotheses



In the following, let us assume that we are interested in finding out which of two theories have more support from evidence (competing theories) across many cases, even though we will limit ourselves in our example to analyzing only one case.

The main steps in congruence analysis as put forward by Blatter and Haverland (2012, p. 167-202) are then:

- selecting the broad theories of interest (e.g. rational choice, self-determination, ...), followed by selecting relevant cases;
- hypotheses are then elaborated for each broad theoretical explanation about how causes are related to the outcome. Note that we are not theorizing a causal mechanism; instead we are formulating factors that can be expected to occur in a temporal fashion if a given theory is valid;
- then relations between hypotheses are considered: are there hypotheses that, if substantiated, automatically contradict a hypothesis in another theory? Do hypotheses overlap, meaning they are present in several broad theories? Do hypotheses complement each other, meaning they are unique to a theory without contradicting those in another theory?
- similarly, to process tracing, a distinction is drawn between the level of theoretical hypotheses and their observable implications for which data can be gathered;
- finally, the relative strength of evidence for the various theories relating to the case(s) at hand is determined.

We elaborate these various steps further below for the PDP.

Selecting theories in congruence analysis

Congruence analysis would start by brainstorming about what it is about an intervention that might be a cause of the outcome. As stated earlier in this article, for the PDP, rational choice theory can provide one plausible explanation. In addition, next to rational choice, we could propose another broad theory, drawing on the idea of vacancy chains, as explained by Hedström (2005), which states that people will move to new positions when these become vacant and that this creates a dynamic throughout the organization (a chain of people filling up vacant positions). The PDP can be theorized to create such opportunities. In Hedström's (2005) framework it is an opportunity based theory rather than one working on knowledge or beliefs.

Elaborating theoretical hypotheses in congruence analysis

We could then propose two different sets (non-exhaustive but used for illustrative purposes only) of theoretical hypotheses. In both cases, these are linked loosely to each other, reflect a time sequence and cover most of the phases of the PDP as depicted in figure 1 and 2:

The rational choice hypotheses could be:

- The PDP project triggers participants to gain more insight in their own competences;
- They also gain more insight in their personal interests and in what they value in work;
- They increase their understanding of possible future career paths;

- Based on such insights, participants will make informed choices regarding the development issues they need to address;
- Participants will draw up action plans that address these identified development issues;
- Participants execute these action plans and acquire or strengthen the necessary competences;
- Participants apply for other jobs or execute their current jobs better.

The same scope conditions already present in figure 1 and 2 are maintained. The vacancy chain hypotheses, on the other hand, could be:

- the PDP project triggers the organization to set up internal mobility processes;
- Highly Motivated Employees (HMEs) will swiftly volunteer to participate in a PDP to take advantage of this opportunity;
- HMEs will move very rapidly (in the PDP) through the reflection stage without having to be coached much;
- HMEs will execute their action plans highly systematically and very quickly;
- HMEs will respond and apply rapidly for new or vacant positions.

Other scope conditions must be introduced to accommodate this explanation, namely regarding the presence of highly motivated employees (HMEs) in the organization who want to advance their careers and already have a good idea of how they want to develop, but who could not proceed due to a lack of structured opportunities. In other words, the research will now apply to a smaller population of cases than the rational choice one.

Elaborating observable implications in congruence analysis

For each of the above theoretical hypotheses, we then need to think of observable implications (empirical predictions) (Blatter and Haverland, 2012, p. 185-7). These can be in the form of single predictions, or a cluster of independent predictions for a given hypothesis that together would have a higher degree of uniqueness (Beach and Pedersen, 2016: p. 286-288). For example, regarding the hypothesis that “HMEs will move very rapidly (in the PDP) through the reflection stage without having to be coached much”, we could expect to observe that records of the meetings show variance in terms of length and number per participant, with HMEs going significantly faster and needing less meetings than non-HMEs. This test is based on regularity across several observations. If there are sufficient numbers, it could be in the form of a statistical test. In addition, we could also ask the coaches how smoothly the meeting went, where we expect that they link HMEs to oddly smooth meetings, at least as compared to non-HMEs. Of course, it is then pre-supposed that we can find out who the HME’s are, e.g. by having talked to supervisors and double checked with the employees. Otherwise, it is not possible to ascertain whether the scope conditions for the opportunity based theory hold for the case.

The type of evidence used in the above examples is “pattern evidence”, as opposed to trace, account or sequence evidence, as discussed by Beach and Pedersen (2013, p. 99-100). If it is not possible to execute a statistical analysis, this is not as such problematic, contrary to what is implied by Beach and Pedersen (2013) who define patterns as statistical in principle. In congruence analysis;

what counts is relative strength of the evidence supporting each theory, hence, even without statistics, it is still possible to say which (set of) hypotheses is more congruent with whatever data is at hand. Representativeness (uniqueness) of the evidence found, will in this approach however also have to be established. This is of course easier with statistical analysis. In addition, certainty of the evidence also should be assessed. Finally, the data should be reliable.

Congruence analysis is very like what Yin (2003: p. 116-7) refers to as "pattern matching", if the "pattern" refers to the overall set of empirical predictions concerning theoretical hypotheses that must be evidenced. This not to be confused with the "pattern evidence" discussed earlier, which refers to a specific kind of evidence.

Determining relative strength of evidence for different theories in congruence analysis

Ultimately, the set of hypotheses needs to be assessed as a whole for each theory using the actual evidence relative to the observable implications, considering also which of the hypotheses are critical for each theory. There can be hypotheses that have a secondary character, meaning they are not strictly 'necessary' for the theory to hold, whereas others are central to the theoretical explanation (e.g. a hypothesis about choice based on a rational evaluation of options is vital to any rational choice explanation).

However, at this point, as we are going for a competing theory approach, we are of course interested in observable implications that, when the evidence is found, are in line with a first theory but in contradiction with another or the reverse. These have the greatest discriminatory power to make a case regarding the support from the evidence for one theory versus another.

Indeed, if we would find that HMEs are NOT having less and shorter meetings, then this would provide some disconfirmation for the vacancy theory (due to high certainty). In addition, we would not expect to find this under the rational choice theory at all, meaning the test exhibits high uniqueness for the opportunity based theory. In fact, it would be in contradiction to the rational choice theory where we expect that meetings take roughly the same amount of time and frequency as the tools to be used should set most of the pace. The expectation that there is little variance constitutes a test with relatively high certainty for rational choice as well as high uniqueness relative to the opportunity based theory.

The example makes clear that multiple theory congruence analysis is implicitly based on the same Bayesian probability calculus already explained in the section on process tracing. As stated by Beach and Pedersen (2016, p. 183), confirmation of a hypothesis in one theory does not necessarily enable us to make any inference concerning the other theory unless two theories have mutually exclusive predictions of evidence and no other theories are considered, as in the example observed. However, congruence analysis does not depend on formulating such contradicting observable implications. Instead, what is important is that tests are formulated for the key hypotheses of each theory, which are then evaluated to see whether there is supporting or disconfirming evidence for them.

Ultimately, to determine the relative strength of theories, the following considerations apply, which go beyond only thinking of contradicting observable implications:

- theoretical hypotheses that are similar (overlapping parts of theories) can obviously not be used to differentiate one theory from another. However, the “weight” of these hypotheses in the theories can be different (critical for one, peripheral for another). So, empirical tests (confirming/disconfirming) may still make a difference when looking at the theories as a whole;
- the hypotheses that are not overlapping can be used for tests with high confirmative power / theoretical uniqueness (smoking guns). However, one should not forget that uniqueness could be low, as the evidence could be equally likely under the alternative theory, even if it is not linked to an explicit core theoretical hypothesis in the alternative theory. What is specific to congruence analysis is that uniqueness is understood with reference to at least one known alternative theoretical hypothesis that can account for it also, which implies that the same test that has high uniqueness relative to only one other theory, can still have low uniqueness in general (when considering all possible alternative theories);
- alternatively, tests for non-overlapping hypotheses that are founded on certainty (disconfirmative power) can be put forward. In this case, there is no reference to the alternative theories: disconfirming stands on its own. That means disconfirming one theory does not have any implications concerning evidential support for the other theory;
- the “quality” of the pattern match at the level of theoretical hypothesis, as substantiated with the actual evidence for the observable implications *h* is of importance. A “pattern” can be 100% as expected but supported by relatively weak evidence. For example, when we have three observations that evidence the pattern exactly as predicted by theory A and a pattern that matches only 80% of what theory B predicts, but with 100, representative, observations (e.g. making statistical analysis possible), then the latter test may “weigh” more in the overall assessment of relative support based on evidence for the theories.

To conclude, if theory A fails more hoop / smoking gun pattern tests on core hypotheses than B, using a holistic understanding of the weight of the particular hypothesis in the whole set of hypotheses for both theories as well as an assessment that the quality of the patterns is equal, then A can be seen as having weaker support from evidence than B.

Such a full consideration of the Bayesian perspective is not present in Blatter and Haverland (2012) but could add value. It would be close to what Befani and Mayne (2014) are advocating in their discussion concerning contribution analysis. It also means that drawing conclusions in congruence analysis can benefit from the procedures for aggregation of evidence that were proposed by Beach and Pedersen (2016). Only for congruence analysis, there is one more level of aggregation, namely where conclusions have to be drawn regarding *which* theory has the most support from evidence.

What would have been the (dis)advantage of using congruence analysis relatively to the original methodology?

While process tracing has received some attention in an impact evaluation context in recent years (as made clear by the earlier references), congruence analysis does not seem to have attracted similar levels of debate. The approach is mentioned in Stern (2015) and Stern et al (2012) but not elaborated clearly in the way conceived by Blatter and Haverland (2012).

Going back to the actual PDP evaluation, it is clear that a fully formed alternative theory was never specified. It therefore is in fact an instance of a single theory congruence analysis. However, one of

the earlier cited pieces of evidence in the PDP study, “There are a number of things. But those we knew already before. They just came up again.”, provides an indication that an opportunity based theory could be worthwhile to research.

Several possible conclusions could then result from a congruence analysis. A first one could be that the opportunity based theory is better (or less) supported by the evidence than the rational choice one, at least within the population of cases with the added scope conditions relating to HMEs. Of course, it could also be the case that rational choice does not receive much support from the evidence regarding what happens with HMEs but neither does the opportunity based one. Clearly, a much richer understanding of what is (not) supported by evidence in the case can result from the approach as compared to sticking to one theory only.

A second advantage is that, as congruence analysis, even in its competing theory variation, does not aim to “rule out” any theory (a “rival”) but merely assess relative evidential support, this approach is less prone to researcher bias than the single theory congruence analysis used to evaluate the case in the original study. Indeed, all theories are assumed to have some merit as the focus is more on assessing relative strength of evidence for each theory, rather than arguing exclusively for a specific theory.

One key challenge when using congruence analysis is that the assessment of the relative strength of different theories can be quite arbitrary. Relative strength is based on the assessment of the centrality of particular hypotheses for theories. If there is disconfirming evidence for a critical hypothesis of one theory, we would downgrade our confidence in this theory in relation to alternatives. However, the centrality of hypotheses is a theoretical question that cannot be tested empirically; it is based on the interpretation of the researcher.

In addition, if observable implications are specified as Bayesian tests, then the same advantages (and disadvantages) this conferred on process tracing apply to congruence analysis.

However, when we are engaging in a competitive theory test, where each theory has multiple hypotheses - some of which are central, others not - and where each hypothesis has multiple observable implications with varying strengths, the complexity of assessing the relative strength of evidence in support of theories can become unmanageable.

Comparing process tracing and congruence analysis

Figure 6 aims to summarize the differences between process tracing and congruence analysis, from which also some relative (dis)advantages can be derived.

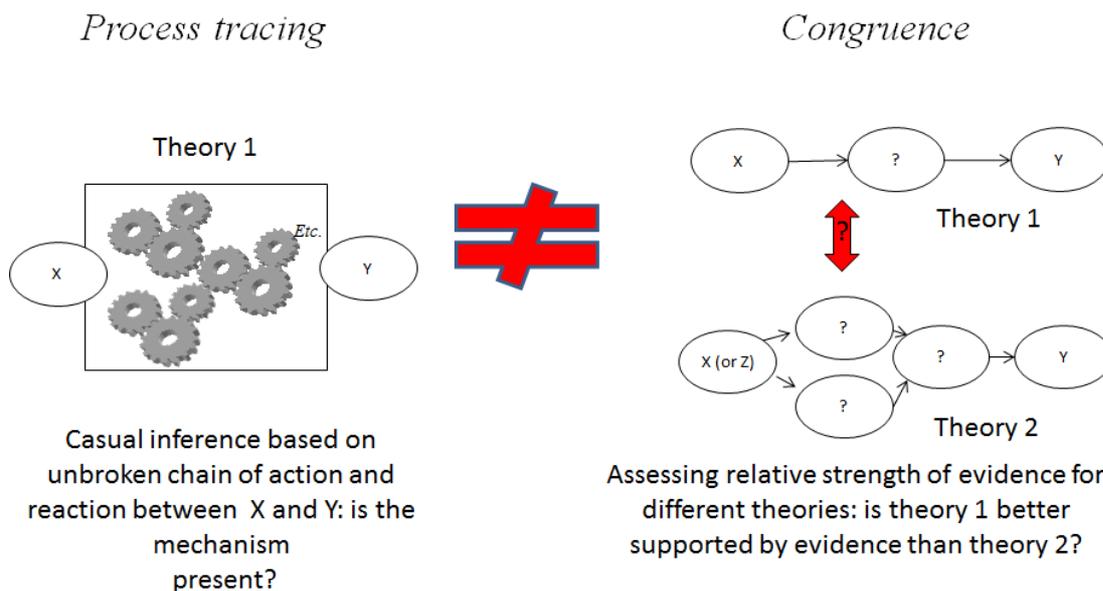
First, process tracing can draw causal inference regarding only one theory, in the guise of a mechanism that is operationalized as an unbroken chain of action and reaction, visualized as the interlocking cogwheels in between cause X and outcome Y. This sheds light on how things work in real world cases. Congruence analysis, in contrast, has only loose, implied connections between a

rich set of hypotheses, constituting at least two theories, whose relations over time can be depicted visually by linking them to each other with arrows. This means conclusions can only be drawn on relative evidential support for each theory, with weaker causal inferences.

However, this also generates the critical methodological advantage of congruence analysis. While it is difficult to use process tracing to test different explanations within a single case due to the amount of analytical resources required, it can become almost impossible to try to do process tracing across multiple cases. In contrast, congruence is more suited to deal with both situations. A congruence analysis can act as a “plausibility probe” that can then focus the analysis in a subsequent process tracing case study on a particular cause and mechanism. If we started with process tracing, we might have inadvertently missed a more important cause/mechanism linked to the outcome. In addition, congruence analysis can be used to be more confident about the scope of generalizations of a theoretical explanation. But, as the number of theories and/or hypotheses grows, this advantage of congruence can turn into a disadvantage as the analysis can become extremely unwieldy.

Finally, in process tracing the assessment of the evidence supporting parts of mechanisms is not dependent on how “critical” the part is expected to be as with hypotheses in multiple theory congruence analysis.

Figure 6: process tracing versus congruence analysis



We want to reinforce the point that the logic of causal inference as used in process tracing should not be applied to congruence analysis. The focus of congruence analysis is not to show that there is a causal link between two propositions, as these links are only loosely specified anyway (no unbroken chain of action and reaction was there to start with). Hence, the visualization with arrows linking propositions can be somewhat misleading; instead what is being depicted is the temporal order in which hypothesized elements occur. In congruence analysis, the focus is rather on using data to ascertain relative support from evidence for at least two fully specified theories (depicted by the red arrow in between the two theories in figure 6).

Conclusion: process tracing or congruence analysis?

When comparing process tracing with congruence analysis we find relative (dis)advantages.

First, congruence analysis, is typically less demanding in terms of theory development, and is significantly easier to implement in practical terms. Process tracing can be expected to require multiple iterations of theory development and testing. However, when the number of theories and associated hypothesis used in congruence analysis grows, the advantage can turn into a disadvantage due to the need to (arbitrarily) state what are the core hypotheses and the complexity of dealing with multiple empirical tests of varying strength.

Second, therefore, process tracing, in order to be realistic, will typically focus on a narrower scope (e.g. only phase 1 of the PDP) than congruence analysis which takes a more high-level approach to hypothesizing (hence the example covered practically the whole PDP process for both explanations).

Third, in terms of specifying empirical tests, with process tracing we typically focus on only one mechanism by asking ourselves if an observable implication is so specific to that theory that it would be relatively improbable that we observe it due to any other theory from a universe of undetermined alternate theories (uniqueness). In congruence analysis, in contrast, we are typically using several

specific and fully specified alternate theories. We leave out of consideration other theories that we did not fully specify. The advantage to this is that it is easier to reflect on how unique an observable implication is for each specific theory, rather than for one theory and an unknown universe of possible theories. However, as the number of theories increases, this advantage disappears.

Finally, the biggest disadvantage for congruence analysis may be that it can only make statements concerning the relative evidential support for a theory, whereas process tracing can establish the presence or absence of a mechanism, enabling stronger causal inferences and shedding light on how things really work.

Evaluators will need to consider these relative (dis)advantages carefully when deciding how best to respond to expectations of stakeholders, given available resources and time constraints. It seems fair to conclude that if stakeholders are interested in the relative merit of a limited number of explanations across a number of cases, then congruence analysis may be a good starting point. However, when there is an interest in making a more definitive judgement regarding one specific explanation for a given case, then process tracing may be the better choice.

References

Astbury B and Leeuw F (2010) Unpacking black boxes: Mechanisms and theory building in evaluation. *American journal of evaluation* 31 (3): 363-81

Bamanyaki P A and Holvoet N (2016) Integrating theory-based evaluation and process tracing in the evaluation of civil society gender budget initiatives. *Evaluation* Vol. 22(1): 72–90

Beach D and Pedersen R B (2013) *Process-tracing methods*. Ann Arbor: University of Michigan Press.

Beach D and Pedersen R B (2016) *Causal case study methods*. Ann Arbor: University of Michigan Press.

Beach D and Pedersen R B (2016b) Selecting appropriate cases when tracing causal mechanisms. *Sociological methods and research*: 1-35

Befani B and Mayne J (2014) Process tracing and contribution analysis: a combined approach to generative causal inference and impact evaluation. *IDS bulletin* 45 (6): 17-36

Befani B, D'Errico S, Booker F and Giuliani A (2016) Clearing the fog: new tools for improving the credibility of claims. IIED briefing April 2016.

Bennett A (2014) Appendix in Bennett A and Checkel JT (eds.) *Process tracing: from metaphor to analytical tool*. Cambridge: CUP.

Blatter J and Haverland M (2012) *Designing case studies*. Houndmills: Palgrave Macmillan.

Cartwright N and Hardie J (2012) *Evidence-based Policy: Doing It Better: A Practical Guide to Predicting If a Policy Will Work for You*. Oxford: Oxford University Press.

Chen H (2005) *Practical program evaluation*. London: Sage Publications.

Chen H (2006) A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, 13 (1): 74-82.

Coryn C L S, Noakes L A, Westine C D, Schröter D C (2011) A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation* 32(2) 199-226

De Rick K, Drieghe L, Jacobs L, De Cuyper P (2014) Het POP: een effectief instrument in het arbeidsmarktbeleid? Een kwalitatieve analyse van de uitvoeringspraktijk en effecten van POPs in drie cases. Onderzoek in opdracht van ESF-Agentschap Vlaanderen.

European Commission (2013) *Evalsed Sourcebook: methods and techniques*. Brussels: European Commission.

Funnel S C and Rogers P J (2011) *Purposeful program theory*. San Fransisco: Jossey-Bass.

Hedström P (2005) *Dissecting the social*. Cambridge: Cambridge University Press.

Khandker S R, Koolwal G B; Samad H A (2010) *Handbook on Impact Evaluation: Quantitative Methods and Practices*. World Bank

Leeuw F L (2012) Linking theory-based evaluation and contribution analysis: three problems and a few solutions. *Evaluation* 18 (3): 348-363

Miles M B, Huberman A M, Saldana J (2014) *Qualitative data analysis*. London: Sage Publications.

Pawson R (2003) Nothing as practical as a good theory. *Evaluation* 9 (4): 471-490

Pawson R and Tilley N (1997) *Realistic Evaluation*. London: Sage publications.

Punton M and Well K (2015) *Straws-in-the-wind, hoops and smoking Guns: What can process tracing offer to impact evaluation?* Centre for Development impact practice paper number 10: April 2015.

Ryan R M and Deci E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78

Schmitt J and Beach D (2015) The contribution of process tracing to theory-based evaluations. *Evaluation*. vol 21 (4) 429-447

Schneider C Q and Wagemann C (2012) *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.

Stern E (2015) *Impact Evaluation: A Guide for Commissioners and Managers*. London: Bond.

Stern E, Stame N, Mayne J, Forss K, Befani B (2012) *Broadening the range of designs and methods for impact evaluation*. London: Department for International Development.

Van Evera S (1997) *Guide to methods for students of political science*. Ithaca: Cornell University Press.

Yin R K (2003) *Case study research design and methods*. Thousand Oaks: Sage.

Yin R K (2013) Validity and generalization in future case study evaluations. *Evaluation* 19 (3): 321-332