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

***EES16-0393***

**Paper presented at the 12th European Evaluation Society Biennial Conference**

**Evaluation Futures in Europe and beyond: Connectivity, Innovation and Use**

**29 September 2016 Maastricht, the Netherlands**

**(Draft 31/8/2016)**

**Benedict Wauters**

**Department of labour and social economy, Ministry of the Flemish Government**

**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

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

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 congruence analysis (as described in Blatter and Haverland, 2012) , including relative advantages and disadvantages. Both process tracing and congruence analysis are analytical strategies that can be used for causal inference within a single case study.

**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 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 labour 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). 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 in order 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.

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 materialise (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

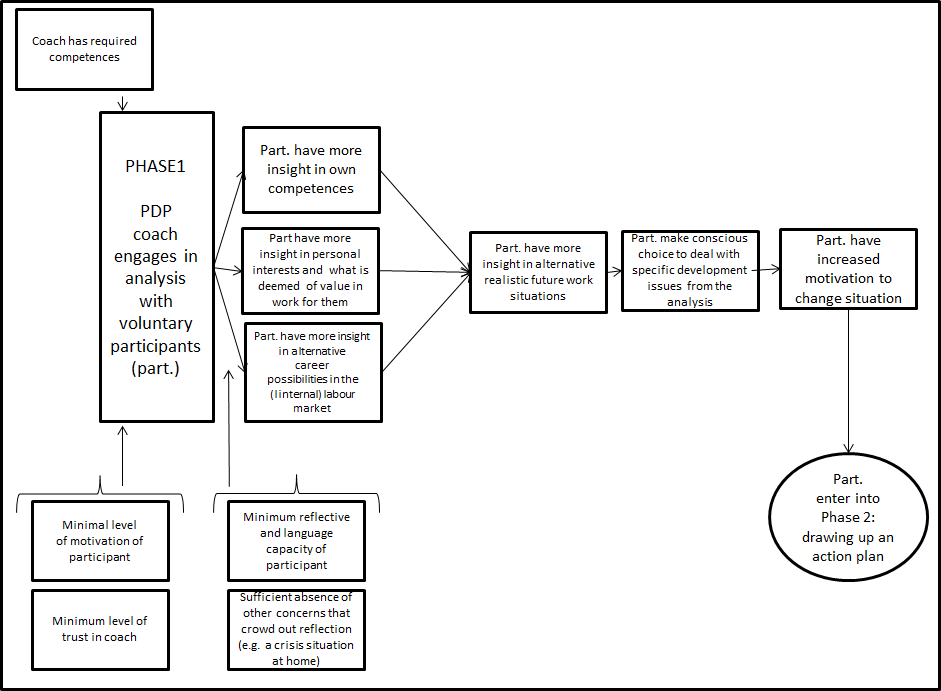
For illustrative purposes, the requirements in terms of the nature of the PDP support are detailed below, but similar descriptions, based on previous exploratory research, were drawn up for all the elements of the action model:

* the whole process should take between two to six conversations with a coach, covering in their totality one and a half to eight hours of engagement.
* it should cover all the parts of the PDP cycle (analysis, option formulation, realization, evaluation)
* there should be differentiation according to the characteristics of the target group (so called vulnerable groups may require more time) and the number of PDP cycles an individual has already gone through (a person who does this for the first time needs more time than someone who is used to such processes).

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

Figure 1 shows what is assumed to occur on the basis of a first (phase 1) coaching intervention with the aim to identify which competences participants should develop in order 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. All of these insights combine into a realistic expectation that they should be motivated to pursue.

Figure 1: detail of phase 1 of the PDP cycle

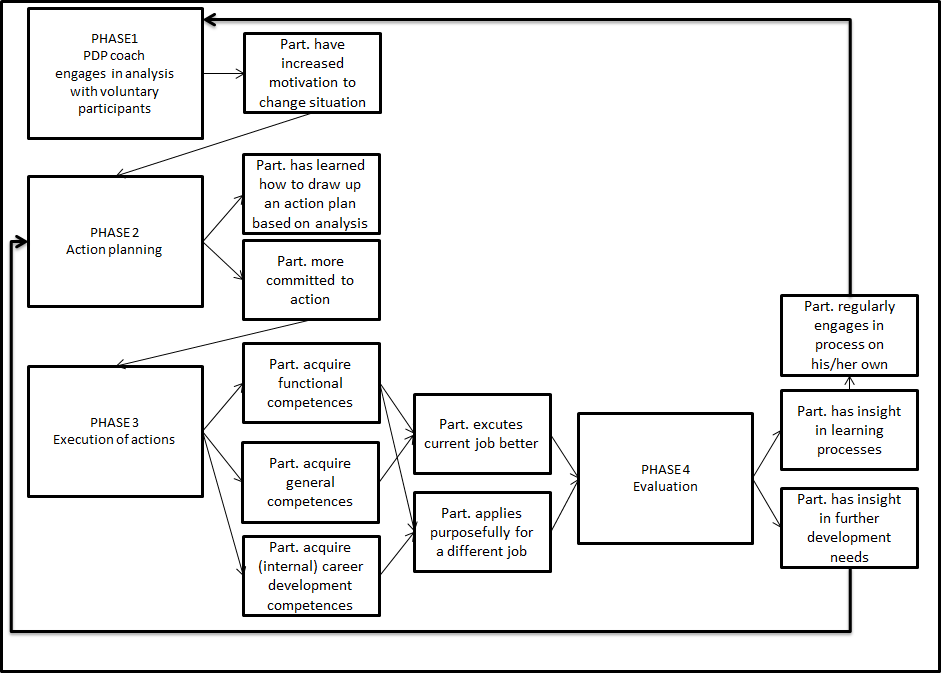


Adapted from De Rick et al (2014)

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 organised 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)[[1]](#footnote-1).

**Findings of the evaluation**

In terms of assessing the fidelity of this case to the action model, the evaluators found that the analysis phase was not very elaborate. This is due to the very restricted scope of development and learning as decided upon in this project which relates only to improving one’s functioning within an existing job or towards a new job that has already been decided beforehand (so not due to the analysis within the PDP). There was no attention to exploring possibilities beyond the current or already decided future position and there is 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 concept is that in this case a demand-led approach, where the participant is central, was not followed. The coach rather steers the participant quite substantially towards the goals of the organization. Phase 1 and 2 (analysis and action plan) are taken together in only one conversation. The execution of the action plan is not supported with more conversations. The evaluation phase (12 months later) is rather part of a new cycle instead of being self-standing. Essentially, in this case, 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 is 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. The evaluators state that, given the restricted understanding of the 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, in three out of the four studied individual PDPs, moderate (due to the restricted scope) levels of insight into competences are noted. In one PDP, only a confirmation of existing insights is noted.

The next sections will use this particular 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. 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 hypothesised outcome or “process whereby causal forces are transmitted through a series of interlocking parts of a mechanism to produce an outcome” (Beach and Pedersen, 2013, p. 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.

For example, Schmitt and Beach (2014) develop such a chain for a part of the reasoning behind budget support in the context of development aid as depicted in Figure 3. 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 states that “pathway from resource to reasoning is referred to as the programme mechanism”. The attempt below is to create an unbroken chain, rather than loosely connected parts.

Figure 3: mechanism as chain of action and reaction in budget support

Source: Schmitt and Beach (2014)

Beach and Pedersen (2013, p.26-7 ) explain that the deterministic foundation of process tracing refers to causality as consisting in necessary and/or sufficient causes. This puts forward that X is a necessary cause of outcome Y when we never find Y without X (NO X, NO Y). X is presumed a sufficient cause of Y if every time we have X, Y also follows (X , thus Y).

But a sufficient cause leaves the possibility of many other X’s that can also cause Y. In many cases we face a contributing or “INUS” condition (Beach and Pedersen, 2013, p.30): an Insufficient but Necessary part of an Unnecessary but Sufficient cause. In process tracing, any of the “steps” in the depicted chain could be seen as INUS. Each step is needed for the chain to function, hence necessary, otherwise the chain as a whole breaks down. But each step is not enough on its own as it requires all the other steps as well. At the same time, the chain as a whole is not necessarily the only route to the outcome: it is not necessary, although it may be sufficient. Whether the whole chain is sufficient or not cannot actually be demonstrated by tracing a single mechanism. It may require multiple mechanisms. Ultimately, sufficiency cannot ever be fully demonstrated.

At this point, it is needed to distinguish between three types of process tracing as elaborated by Beach and Pedersen (2013, p. 56-67):

* Theory testing process tracing: here the aim is to research whether a mechanism of interest is indeed present and accounts for at least part of the observed outcome of a case. However, the interest is not in fully explaining the outcome. Many other mechanisms might be needed for that. Usually, there is an interest in examining multiple cases to ascertain if a mechanism is unique to a case or travels across cases;
* Outcome explaining process tracing: here the aim is 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. There usually is no interest in generalizing across cases;
* Theory building process tracing: here the aim is to hypothesise the process that may link the cause to its outcome when one does not yet have a theory about this.

In impact evaluation, the aim is 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 causal processes, ready for testing. This could be part of “evaluability” study to be conducted before the actual impact evaluation. In many cases, such a prior study will be required.

In both theory testing and outcome explaining process tracing, theorizing a mechanism as an unbroken chain of action and reaction, combined with deterministic causality, means that if any step in the chain can be falsified on the basis of evidence, then the cause X cannot have led to the outcome Y. By falsifying part of the mechanism, the whole mechanism is falsified.

The difference between a theory testing approach an and outcome explaining one is that in the latter case, we will have to conduct this procedure for more mechanisms, until we are reasonably satisfied they explain most of the observed outcome. In contrast, a theory testing approach attempts to generalize across cases, hence the same mechanism will have to be researched in multiple cases.

Generally speaking, an outcome explaining approach will represent a more substantial piece of work as it implies a much broader understanding of theoretical mechanisms, all of which need to be conceptualised as processes. The theory testing approach requires understanding of only one mechanism, which, initially, can be researched in only one case. Of course, it then is not possible to generalize beyond that case but that may not be immediately 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 a number of 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

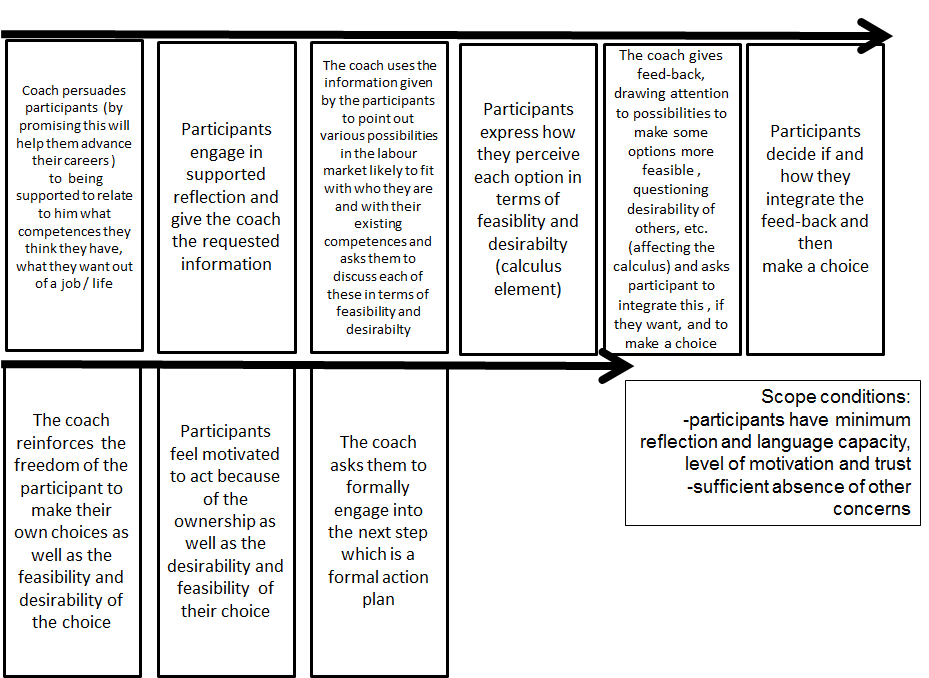
3) Substantiating the presence of the mechanism by proposing observable implications for all the steps in the mechanism and gathering the corresponding data.

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 actors. Figure 4 shows what the PDP theory of change (comprising the action and change model) for phase 1 could look like as a mechanism.

Figure 4: 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 provide for the necessary causal force. 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). This is in line with the original change model in Figure 1 where insights lead to choices being made. However, the mechanism display requires to make exactly cleat how these insights lead to this choice being made, within the overall rational choice assumption.

**Selecting cases in process tracing**

Beach and Pedersen (2013, p. 150) state that only cases that have achieved a certain threshold of cause X as well as the outcome Y should be selected if the interest is to trace the process from X to outcome Y. However 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) has been reasonably faithfully implemented as well as achieved the desired (intermediate) outcome Y (in the PDP example, 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.

A first important step in process tracing must therefore be to identify the population of cases and to find out whether X and Y are (sufficiently) present or not. Let us assume that the original phase 1 action and change models had been respected and that there are indeed participants that decide to got to phase 2. Hence the rational choice mechanism as described in Figure 4 can reasonably be hypothesized to apply. How would we proceed then to substantiate this mechanism?

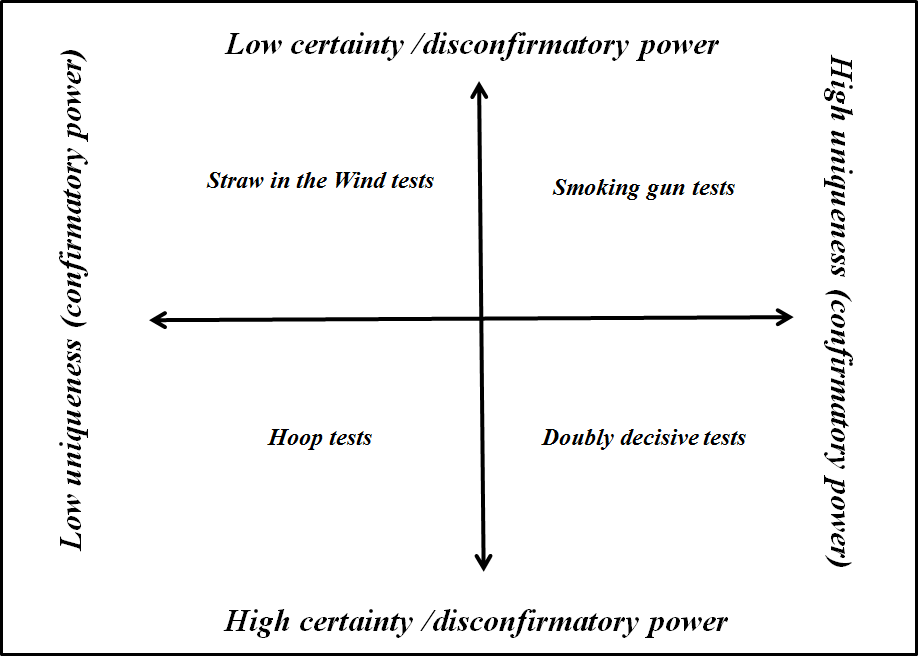
**Substantiating the presence of the mechanism in process tracing**

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 a 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 4). 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 presumably have to be gathered via interviews with coaches and/or participants.

However, the above test then begs the question what kind of test of the presence/absence of the theoretical mechanism isactually being proposed? Beach and Pedersen (2013) provide a powerful way to proceed, depicted in Figure 5, based on a framework by Van Evera (1997), which is underpinned by Bayesian probability calculus. For detailed explanations about Bayesian probability calculus we refer to Beach and Pedersen (2013, p. 96-99).

Figure 5: typology of empirical tests



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.
* Certainty: 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

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 in a report could have been derived in many different 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 falsified. 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.

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 in terms of their competences that they did not have before”*. 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 generated 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 and 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).

Finally, 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.

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

Schmitt and Beach (2014) 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;
* A clearer logic of inference (Bayesian tests);

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 positions 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 prioritise 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 particular cause was relative to other causes. In addition, process tracing requires the outcome of a case to be fully known and can therefore be conducted only at a moment in time when this information is available. 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 particular 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 (2014) 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 more or less 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.

Beach and Pedersen (2013, p. 38) state that a theory as depicted in 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 commitment to 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 speak 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.

Next, in line with Schmitt and Beach’s (2014) second advantage, it is clear that using the Baysian empirical test framework, a greater number of tests have been devised than are present in the original evaluation. In addition, it would be more transparent what the value of these tests is.

If we look at the actual research conducted by the original evaluation, the conclusion that insights 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 actually does not tell us anything: insights could have been derived in many different 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 disadvantages of process tracing, we can add an additional three to those already specified by other authors:

* 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 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 in order to specify case specific empirical tests;
* Finally, the approach is 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 theorised process, followed by more testing of all 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 only.

Both disadvantages 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 (2013, p. 28), based on Hume’s idea of causality as constant conjunction (same effects appear with regularity after the same cause) as well as determinism. In addition, contrary to mechanistic thinking which drew on an unbroken chain of cause and effect to substantiate causality, congruence analysis draws on the idea of the relative power of different theories to explain an outcome. Indeed, in the absence of an unbroken chain of cause and effect between X and Y, there is no way to make causal inference from X to Y. No matter how many times we observe a deterministic regularity of X thus Y, hence pointing towards X as a sufficient cause, we can never rule out that both Y and X appear because of a cause Z that is sufficient for both Y and X, without there being any direct causal link between Y and X. This the same issue that has given rise to counterfactual impact evaluation methods (see e.g. Khandker et al, 2010) although traditionally there causes are not viewed deterministically but probabilistically (in terms of tendencies rather than sufficiency or necessity).

However, what remains possible, barring a counterfactual, is to compare different theories that embody different ways of X being sufficient for Y (there is more than one road from Brussels to Rome). Alternatively, theories can also be selected that seek to embody different Z’s being sufficient for Y (there are many more cities than Brussels that can lead to Rome). It should be noted therefore that congruence analysis is not attempting to definitively “rule out” alternative theories, but rather to establish comparative strength (what road seems to be the more travelled). 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, they also state that this is a weak alternative to using multiple theories.

Just like in process tracing there could be an interest in researching if a particular explanation travels across cases. Then we could be interested to understand which of several theories would have the most strength in explaining a range of cases (referred to as the competing theory approach, in Blatter and Haverland, 2012, p. 145 ). However, similarly to the case centric, outcome explaining process tracing that is oriented to explaining a case sufficiently, we could also use different theories in congruence analysis to come up with a reasonably sufficient explanation of the outcomes of a case (referred to as complementing theory approach, in Blatter and Haverland, 2012, p; 145). As with the process tracing example, let us assume that we are interested in finding out which of two theories have more explanatory strength (competing theories) across a number of cases, even though we limit ourselves initially 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 theories of interest, followed by relevant cases;
* next the theoretical propositions are elaborated for each theory. Theory for Blatter and Haverland (2012) consists of propositions that define causal relations between core concepts (the latter are themselves part of larger paradigms) and which, to become more specific theories, are allied to peripheral concepts. So it is possible to have conflict theory, composed of core concepts, but to have specific applications of that theory within the field of families versus organizations (hence adding peripheral concepts) ;
* then relations between these propositions are considered: are there propositions that, if substantiated, automatically contradict a proposition in another theory? Do propositions overlap, meaning they are present in several theories? Do propositions complement each other, meaning they are unique to a theory without contradicting propositions in another theory?
* similarly to process tracing, a distinction is drawn between the level of theoretical propositions and their observable implications for which data can be gathered;
* finally the relative explanatory strength of 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**

As stated earlier in this article, for the PDP, rational choice theory can deliver the core concepts of interest. In addition, next to rational choice, we could propose another theory, drawing on the idea of vacancy chains, as explained by Hedström (2005), which states that people will move to new positions if and 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 as it requires that coaches discuss realistic future work situations, based on a good view of what is possible/desirable within the organization, including in terms of new positions. In Hedström’s (2005) framework it is an opportunity based theory rather than one working on knowledge or beliefs.

**Elaborating theoretical propositions in congruence analysis**

We could then propose the following (non-exhaustive but used for illustrative purposes only) theoretical propositions. These are linked loosely to each other and cover most of the phases of the PDP:

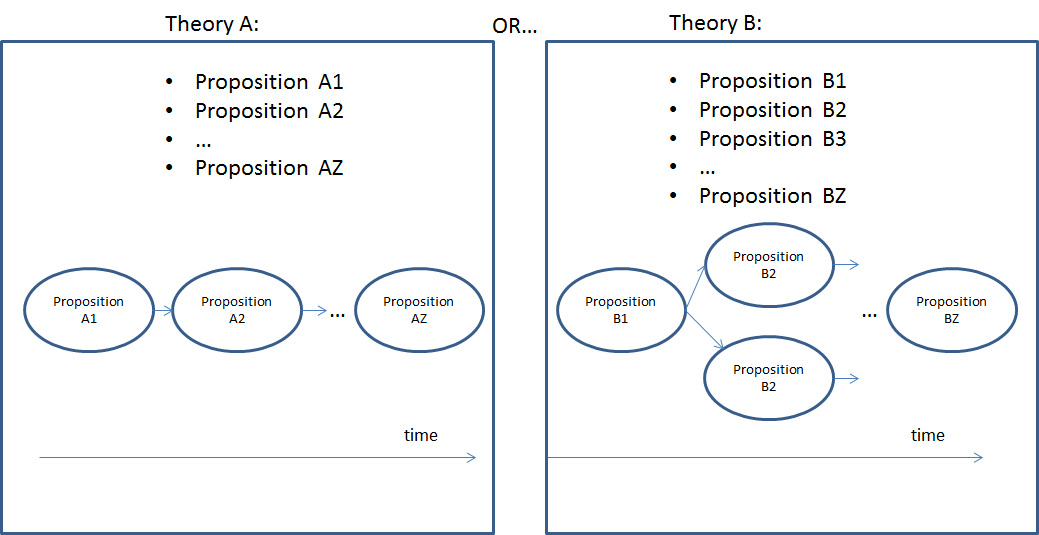
* highly motivated employees (HMEs) in the organization who want to advance their careers and already have a good idea of how they want to develop, could not proceed due to a lack of structured opportunities;
* the PDP triggers the organization to set up internal mobility processes;
* HMEs will swiftly 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 more systematically and faster than other employees;
* HMEs will respond and apply more rapidly for new or vacant positions.

The rational choice propositions could be:

* Participants in the PDP gain more insight in their own competences;
* They 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 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.

It should be noted that Blatter and Haverland (2012, p. 163-4) do not require to visualize the theory as in Figure 1 but rather formulate it as a rich set of coherent propositions that provides a rich enough characterization of the theory. This must be done for all theories one wishes to use. There relation between propositions and their visualization is depicted below. The visualization of how propositions are linked to each over time can clearly add value. However, it should be understood that this is not similar to process tracing where a mechanisms is depicted as an unbroken chain of action and reaction.

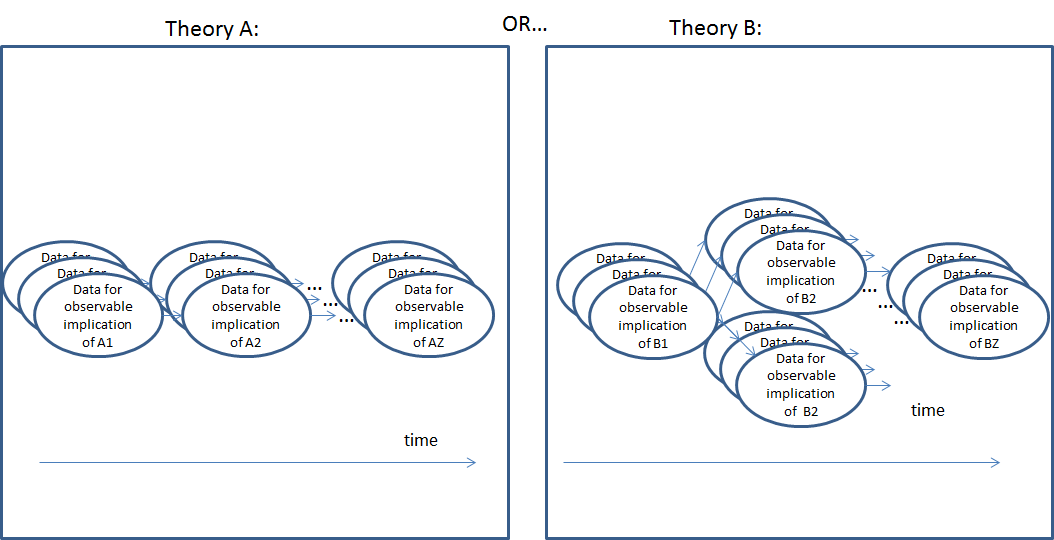
Figure 6: a theory as composed of a rich set of loosely linked propositions



**Elaborating observable implications in congruence analysis**

For each of the above theoretical propositions, we then need to think of observable implications (empirical predictions) (Blatter and Haverland, 2011, p. 185-7). For example, regarding the proposition 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. As it is a regularity based approach, we would indeed expect to find similar findings for individual HME versus non-HME PDPs within the overall case. We could also ask the coaches how smoothly the meeting went. We expect that a coach links HMEs to oddly smooth meetings with regularity, at least as compared to non-HMEs. This is very similar to what Yin (2003: page 116-7) refers to as ”pattern analysis”. For two theories (A and B) this is visualized in Figure 7.

Figure : pattern analysis



The figure makes clear that we need multiple actual observations regarding every single observable implication of each theoretical proposition for each theory.

Of course, in the PDP example, it is 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. Finally, the data should be reliable.

**Determining relative explanatory strength of theories in congruence analysis**

At this point, as we are going for a competing theory approach, we are of course most interested in propositions that 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 relative strength of one theory versus another.

Ultimately, the whole set of propositions needs to be assessed for each theory using the actual evidence relative to the observable implications, taking into account also which of the concepts are really central to each theory. But contradictory observable implications will carry quite some weight in the relative assessment of theories. Indeed, if we would find that HMEs are not having less and shorter meetings, then this would provide some disconfirmation for the vacancy theory. However, if we would find this evidence, this then 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 example also makes clear that multiple theory congruence analysis is implicitly based on the same Bayesian probability calculus already explained in the section on process tracing. Contradicting observable implications as used in a competing theories approach imply an assessment of uniqueness where the observable implication provides confirmation for theory A but not for theory B. This means that conclusions from congruence analysis say nothing on a theory in the absolute but only refer to relative strength or additional explanatory power of the considered theories.

With a contradicting observable implication, certainty also comes into the picture: the evidence for an observable implication that, if present, would rather confirm theory A, must be absent for theory B, or B would be disconfirmed. Of course, if the observable implication must also be present for theory A or it would be disconfirmed, then this is a double decisive test for A. But this does not follow automatically from the concept of a contradicting observable implication. 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.

**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 very much 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 a vacancy chain theory could be worthwhile to research. Several possible conclusions could ensue. A first one could be that the vacancy chain theory has more explanatory strength for the overall case outcome than the rational choice one, because most of the results are linked to HMEs and the vacancy chain theory can be demonstrated to be more congruent for them than rational choice. Of course, it could also be the case that rational choice does not explain very well what happens with non-HMEs and neither does the vacancy chain (which only applies to HMEs), etc. Clearly, a much richer understanding of what is (not) going on in the case can result from the approach.

A second advantage is that, as congruence analysis, even in its competing theory variation, does not aim to rule out a theory, this approach is less prone to researcher bias. Indeed, all theories can have some merit as the focus is more on explaining, rather than arguing for a specific theory.

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

In terms of time and resource requirement, congruence analysis requires a broad theoretical background to be able to specify more than one theory as well as intensive data collection to satisfy the principle of regularity of data.

Finally, congruence analysis can only establish relative causal statements, not absolute ones.

**Comparing process tracing and congruence analysis**

Figure 8 aims to summarise the differences between process tracing and congruence analysis.

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. Congruence analysis, in contrast, has only loose, implied connections between a rich set of propositions, constituting at least two theories, whose relations over time can be depicted visually by linking them to each other with arrows.

Second, where process tracing does not require multiple observations to test observable implications (we only want to demonstrate the mechanism can indeed work in the case at hand), this does not apply to congruence analysis. For the latter, the relative regularity of observations relating to observable implication is a key element in judging which theory has the greater explanatory weight. Next to this, also the relative coverage of propositions by successful (they confirm or disconfirm) tests in the various theories is of importance. However, if theory A contains 11 propositions, 9 of which that are at the core of the theory, which are successfully tested with data, but with little regularity, while theory B contains 10 propositions, 7 of which are successfully tested with great regularity, then we are likely to judge theory B as the stronger one.

Figure 8: process tracing versus congruence analysis

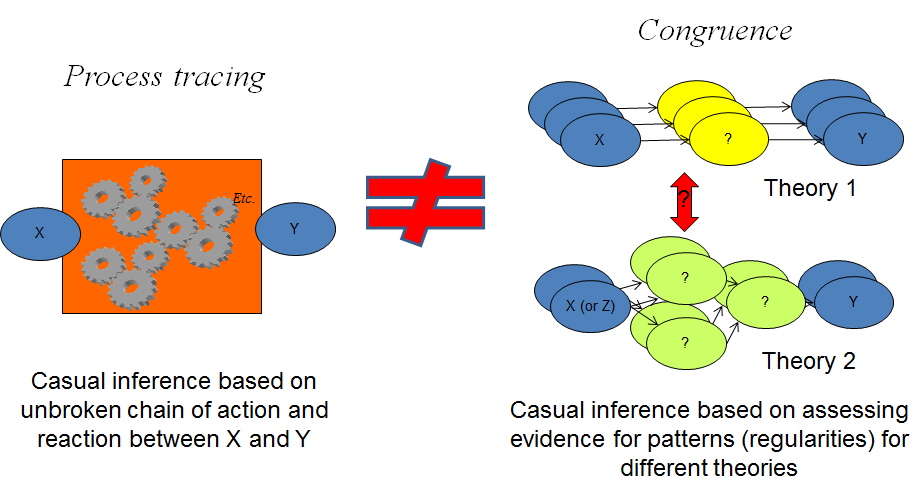
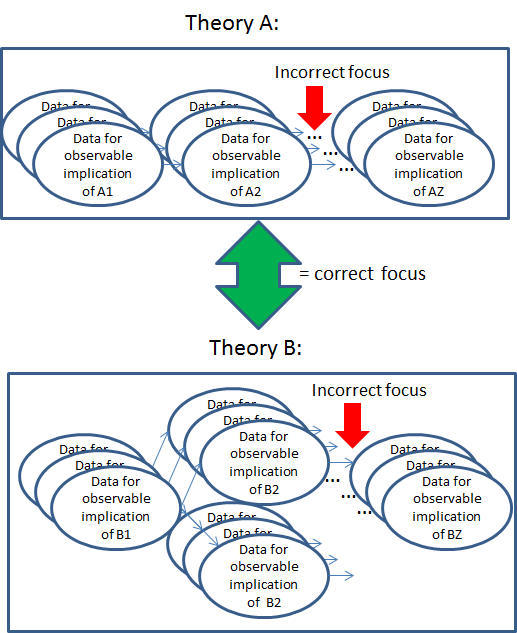


Figure 9 also attempts to reinforce the point that the logic of causal inferencing as used in process tracing should not be applied to congruence analysis. The focus of congruence analysis is not to show that a link between two ensuing propositions does not exists (as in process tracing) , as these links 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 misleading. In congruence analysis, the focus is rather on using data to ascertain relative explanatory strength of at least two fully specified theories.

Figure 9: the focus of congruence analysis



It should be noted that, due to this different focus, unlike in process tracing, congruence analysis can also be used to derive conclusions in situations where an intervention was not implemented to a sufficient degree. Various theories that could explain (part of) the outcomes may be relevant anyway and it would at least be consistent with expectations to find that other theories explain the outcomes better than the one that relies most on an adequate implementation of the intervention.

**Conclusion: process tracing or congruence analysis?**

When comparing process tracing with congruence analysis we find relative (dis)advantages concerning theory development, potential scope, case selection, specification of observable implications, data collection and the nature of the conclusions:

First, congruence analysis, even though it requires elaboration of at least two theories, is less demanding in terms of theory development as it is less sensitive to misspecification of propositions. Process tracing can be expected to require multiple iterations of theory development and testing.

Second, as a consequence, process tracing, in order to be realistic, will typically focus on a more narrow scope (e.g. only phase 1 of the PDP) than, congruence analysis (which covered practically the whole PDP).

Third, in terms of case selection, congruence analysis has the advantage that it can be used with cases that have inadequate implementation of an intervention and still shed light on what is (not) going on from multiple theoretical perspectives.

Fourth, in terms of specifying empirical tests, with process tracing we 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 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 is easier to reflect on how unique and certain an observable implication is for each specific theory, rather than for one theory and an unknown universe of possible theories.

Fifth, congruence analysis is more demanding in terms of data collection as various observations are required to test the same observable implication in order to establish sufficient regularity.

Finally, the biggest disadvantage for congruence analysis may be that it can only make relative statements concerning the explanatory strength of a theory, whereas process tracing can establish the presence or absence of a mechanism.

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.

**References**

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

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

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.

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

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.

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.

Schmitt J and Beach D (2014) Tracing mechanisms as a means to evaluate the governance effectiveness of budget support. Conference paper at 14th EADI general conference, June.

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

1. 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. [↑](#footnote-ref-1)