# Technical Report*

### Assessing a Formal Model of Reflective Equilibrium

Andreas Freivogel, Sebastian Cacean

University of Bern, KIT

3/3/23

Beisbart, Betz, and Brun (2021) introduced a formal model of reflective equilibrium based on the theory of dialectical structures (Betz 2010, 2013), which, according to them, can be used to better understand the method of reflective equilibrium and to assess its potential to yield better epistemic states. However, their discussion was based on a few illustrative examples only, without assessing how the model behaves under a wider spectrum of circumstances.

This document summarizes findings of assessing the RE model more thouroughly by, first, running the model to determine fixed points of the RE process and calculating global optima under a wide range of configurations and, second, by tweaking the original model. The simulation outcomes of three model variants are compared to the ones of the original model.[1]

In particular, we evaluate how the four models perform to various desiderata according to the following subquestions, which we address in individual explorative studies (see below).

- A) Are global optima reachable, i.e. are they fixed points of RE processes?
- B) Are global optima/fixed points full RE states?
- C) Are commitments from global optima/fixed points free of inconsistencies?
- D) Do global optima/fixed points have extreme values (maxima, minima) in their measures?

*Earlier versions of this report were discussed on several occasions with all members of the research project 'How far does Reflective Equilibrium Take us? Investigating the Power of a Philosophical Method' (SNSF grant 182854 and German Research Foundation grant 412679086). We thank, in particular, Claus Beisbart, Gregor Betz, Georg Brun, Alexander Koch and Richard Lohse for their helpful comments, which helped to improve this report considerably.

[1]The results of Beisbart, Betz, and Brun (2021) are based on a Mathematica implementation of the model (see https://github.com/debatelab/remoma). Here, we rely on a reimplementation in Python (LINK TO PUBLIC REPO).

Contact: andreas.freivogel@unibe.ch,

Minimal models should answer the subquestions in the affirmative to a sufficient degree. However, undesirable behaviour of a model variant should be explainable in terms of additional input, apart from the model specification, which is required to run simulations.

Additionally, we address the following questions:

- E) Is the union of commitments and theory of a global optimum/fixed point consistent?
- F) How many steps does an RE process take?
- G) Does an RE promote principles in theories?
- H) Are the results from the published paper reproducible?

## Methods

### Model variants

The *achievement function* plays an important role for both global optima and adjustment steps in RE processes. Global optima maximize the achievement function and candidate commitments or theories in adjustment steps are selected according to their score in the achievement function. Achievement is defined as follows for commitments $C$, a theory $T$, and initial commitments $C_0$:

$$Z(C,T|C_0) = \alpha_A \cdot A(C,T) + \alpha_S \cdot S(T) + \alpha_F \cdot A(C|C_0)$$

where $A, S, F$ are measures for account, systematicity and faithfulness with respective weights $\alpha_A, \alpha_S, \alpha_F$. The measures for account, systematicity and faithfulness are based on a monotonically decreasing function $G(x) = 1 - x^2$ (see Beisbart, Betz, and Brun 2021 for details).

The ensemble studies include four variants of the RE model resulting from a combination of two revisions in the achievement function. First, the monotonically decreasing function `$G$` involves a quadratic term (default in the published paper, but not explicitly motivated) that could be replaced by a linear term. Second, the measure for systematicity involves the ratio between a theory's size and the size of its closure (default in the published paper). A more puristic variant, which does not involve the closure of a theory (its "scope"), relates the size of a theory with the size of the sentence pool.

Combining the model revisions results in four model variants, `QuadraticDefaultSystematicityRE` (in short, `QDS`), `QuadraticPureSystematicityRE` (in short, `QPS`), `LinearDefaultSystematicityRE` (in short, LDS) and LinearPureSystematicityRE (in short, LPS), which are defined as follows:[2]

[2]For these models, systematicity $S$ is maximized by singleton theories, i.e. sets that contain exactly one sentence. This might be considered problematic for the default model variants (QDS,LDS), as they do not discriminate singleton theories on the basis of their scope. Model variants with a different behaviour are, however, not included in this explorative study.

| | $S(T) = G(\frac{\lvert T\rvert-1}{\lvert \overline{T}\rvert})$ | $S(T) = G(\frac{\lvert T\rvert-1}{n})$ |
|---|---|---|
| $G(x) = 1 - x^2$ | QDS | QPS |
| $G(x) = 1 - x$ | LDS | LPS |

### Ensembles

Every simulation of an RE process and the calculation of global optima requires to specify inputs: i) the model variant, ii) the dialectical structure, iii) weighting of the measures, and iv) the initial commitments. Let us call a specification of inputs i)-iv), which allows to run simulation, a *configuration*. Every configuration yields a simulation of a possibly branching RE processess and a set of global optima (i.e. commitment-theory-pairs that maximize the achievement function). Generating an ensemble includes the following tasks and requires the specification of additional parameters.

*Size of sentence pool*

Due to the exponential growth of candidate commitments and theories, which all have to be considered for global optima and semi-global adjustment steps in RE processes, the ensembles include sentence pools with a small number of unnegated sentences (around 5 to 8 sentences).

*Automatically generated dialectical structures*

The creation of random dialectical structures includes the following parameters: The number of unnegated sentences in the sentence pool, the number of arguments, the (maximal) number of premises per argument and whether there is variation, and the (minimal) number of principles. A sentence is called *principle* if and only if it occurs in at least one argument as premise and it or its negation does not occur as a conclusion in any argument. The random generation ensures that the dialectical structure contains the minimal number of principles, but there may be more sentences that satisfy the condition, too.

In addition, the creation of arguments ensures that all (unnegated) sentences are used in the arguments, and there is some preference for sentences that occur fewer times. Before a new candidate argument is added to the other arguments of the dialectical structure under construction, it has to pass a series of test conditions: The new argument - is not question-begging, i.e. the conclusion is not part of the premises - is not attack-reflexive, i.e. the negation of the conclusion is not part of the premises - has premises that are not a subset of premises of another argument - is jointly satisfiable with the other arguments - reduces the number of complete consistent postions, if it is added to the dialectical structure

Apart from generating ensembles with randomly generated dialectical structures, other ensembles are based on handwritten examples (Semmelweiss example, variations of the standard example), that allow for a more intuitive interpretation of results.

The *inferential density* of a dialectical structure $\tau$, $D(\tau) = \frac{n - lg(\sigma)}{n}$, where $n$ is the number of unnegated sentences in the sentence pool and $\sigma$ is the number of complete and consistent extensions in $\tau$, was kept between 0.05 and 0.5.

*Random selection of initial commitments*

There are $3^n$ minimally consistent set of sentences that serve as initial commitments. To create intial commitments by chance, an integer between 1 and $3^n$ is randomly chosen, converted to a ternary representation and initialized as a position in the Python implementation.

*Resolution of weights*

In order to explore different values for $\alpha_A$, $\alpha_S$ and $\alpha_F$ in the achievement function, a resolution (i.e. the separation of the unit intervall $[0, 1]$ into intervalls of equal length) of weights can be specified for an ensemble. In practice, every combination of $\alpha_A$ and $\alpha_S$ are , and $\alpha_F$ is set so that they satisfy the boundary condition $\alpha_A + \alpha_S + \alpha_F = 1$.

The extreme value of $\alpha_F = 0$ is excluded (!from the later ensembles!) since it produces all and only singleton theories (and their closure) to be global optima.

*Overview of ensembles*

This study is based on different ensembles, which differ with respect to the following properties:

| ensemble | rows | dialectical structure generation | dialectical structures | initial commitments | alpha resolution (count) | branching processes | principles |
|---|---|---|---|---|---|---|---|
| 01 | 54000 | old random | 30 | 10 | steps: 0.1 steps (45) | no | no |
| 02 | 45000 (incomplete) | old random | 40 | 10 | steps: 0.1 (45) | no | no |
| 03 | 320 | standard variations | 8 | 10 | a: 0.35, s: 0.55, f: 0.1 (1) | no | no |
| 04 | 3200 (+ 320) | standard variations | 8 | 10 | steps: 0.2 (10) | no | no |
| 05 | 69000 | new random | 39 | 10 | steps: 0.1 (45) | no | no |

| ensemble | rows | dialectical structure generation | dialectical structures | initial commitments | alpha resolution (count) | branching processes | principles |
|---|---|---|---|---|---|---|---|
| 06 | 90000 | new random | 150 | 50 | (0.35, 0.55, 0.1), (0.5, 0.5, 0.1), (0.55, 0.35, 0.1) (3) | no | 2 |
| 07 | 129600 | new random | 60 | 15 | steps: 0.1 (36) | yes | no |
| 08 | 30240 | standard variations | 3 | 70 | steps: 0.1 (36) | yes | no |
| 09 | 314784 | standard example | 1 | 2186 | steps: 0.1 (36) | yes | no |

**Remarks for the interpretation or explanation**

Due to the wide range of required inputs for a configuration, salient behaviour may depend on the features of any of i)-iv) or a combination thereof, and it is difficult to separate all dimension at once. In order to disentangle i) model variants and iii) the weightings, heatmaps proved to be a useful tool and provided insightful plots with respect to various evaluation criteria. The other inputs involve much more features, which could explain salient behaviour, but are much harder to separate: - ii) features of dialectical structures include the size of the sentence pool, number of arguments, number of premises per argument, inferential density, number of complete consistent extensions, number of tau-truths. - iv) features of initial commitments include their size, dialectical consistency, dialectical closure, number of complete consistent extensions, the minimal axiomatization

## General Results

In general, we did not find conclusive evidence to exclude a model variant as a future reference point. Each model variant satisfies the consolidation criteria to a sufficient degree for a wide range of configurations. At this point, undesirable behaviour (failing with respect to some criterion) cannot be attributed to a model variant on its own, but may result from a combination of other aspects of a configuration, such as unfavourable dialectical structures, extreme weightings or hopelessly absurd initial commitments. This does not mean, that there are no differences at all or no tendencies in favour of some model variants. Linear models tend to perform better as quadratic variants with respect to many criteria.

If conceptual clarity is to be taken into account, pure systematicity (number of a theories sentences normalized by the size of the sentence pool) may be preferred to default systematicity, which involves normalization by the size of a theory's closure (its "scope"). In the former case, pure systematicity may safely be called "simplicity" and other virtues contributing to the systematicityof a theory (e.g. scope) should be reserved for model extensions. In this line of thought, `LinearPureSystematicityRE` can be seen as the absolute minimal model variant, that is extended by the other variants by squaring measures, refining systematicity, or both.

## Detailed explorations

This part presents the findings of analysing ensembles 7-9 and evaluating the different models according to the described desiderata.

### A) Global optima and fixed points

#### Background

The formal model of RE allows to distinguish to outcomes of RE, i.e. a theory-commitment-pair for some given intial commitments: *global optima* accdording to the achievement functions and *fixed points*, which are reached by an RE process. Global optima are calculated over theory-commitment-pairs and fixed points result from stepwise, mutual adjustment of commitments and theory. Consequently, global optima and fixed points may fall apart for a configuration. A global optima may not be reachable from given inital commitments by an RE process, or a fixed point may not be globally optimal.

Having a substantial overlap of global optima and fixed points is a desirable feature of an RE model. Otherwise, something would be amiss in the achievement function or in the RE process, for which it provides guidance.

#### Method

Obviously, the data which is required to explore the overlap, consists of all global optima and fixed point for a configuration. The brute force search for global optima requires to calculate the achievement function for every theory-commitment-pair (number of dialectically consistent positions * number of minimally consistent positions, worst case: $3^n \cdot 3^n$ ). However, there are some heuristics that, on average, speed up the calculations dramatically. In order to determine all fixed points for a configuration, all "branches" of an RE process need to be tracked, that occur due to the random choice among equally good performing candidates in an adjustment step. After the construction of global optima and fixed points, one can check for every element whether is contained in the other set.

The overlap for a configuration is given by the the number of global optima, which are fixed points. Note that the number of global optima which are fixed points is equal to the number of fixed points, which are global optima (see the Venn-diagramms in Figure 1). The relative overlap w.r.t. global optima (fixed points) is the ratio of the overlap and the number of global optima (fixed points) for a configuration. In gereral the relative overlaps w.r.t global optima or fixed points,respectively, are not the same, since configurations yield more global optima than fixed points.

**Results**

Observations:

- For every model variant the total number of global optima always exceeds the total number of fixed points between roughly 15% (QDS, ensemble 08) and 50% (QPS, ensemble 07).
- Quadratic models tend to produce more global optima and fixed points than linear variants. (The exception is QDS, ensemble 08)

Mean relative number of fixed points that are global optima

| | ensemble 07 | ensemble 08 | ensemble 09 |
|---|---|---|---|
| LDS | 71.9% | 72.2% | 76.9% |
| LPS | 77.9% | 72.5% | 76.5% |
| QDS | 73.0% | 74.8% | 75.6% |
| QPS | 81.9% | 73.5% | 73.6% |

Mean relative number of global optima that are fixed points

| | ensemble 07 | ensemble 08 | ensemble 09 |
|---|---|---|---|
| LDS | 61.6% | 68.2% | 68.9% |
| LPS | 61.3% | 65.6% | 66.5% |
| QDS | 63.6% | 68.9% | 67.1% |
| QPS | 59.7% | 63.4% | 62.6% |

**Conclusion**

Global optima and fixed point overlap to a sufficient degree in every model variant. Roughly 3 out of 4 fixed points are global optima, and 2 out 3 global optima are fixed points.

The questions of why there are generally more global optima than fixed points, and when they fall apart require more detailed analysis of additional inputs. For example, there seems to be correlation of inferential density of the dialectical structure and the number of global optima or fixed points.

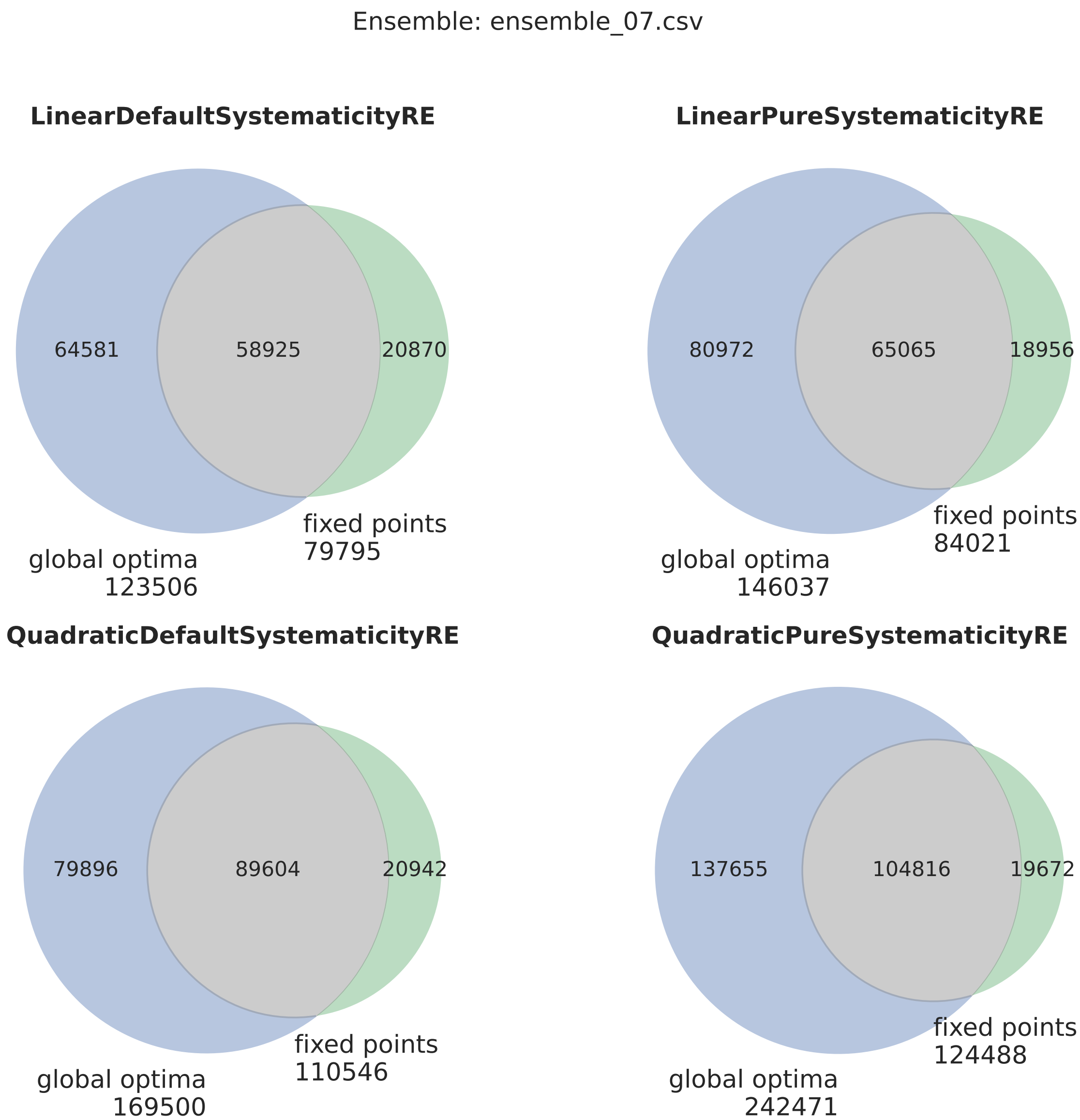


Figure 1: Share of global optima and fixed points.

## B) Full RE states

### Background

A theory-commitment-pair $(C, T)$ is a *full RE state* iff i) it is a global optimum according to the achievement function, and ii) the theory $T$ fully and exlusively accounts for the commitments $C$.

### Method

Formally and irrespecitve of the model variant, full and exclusive account means $C = \overline{T}$, or equivalenty, $A(C, T) = 1$. During the generation of an ensemble, we can store for every global optima and for every fixed point resulting from a configuration, whether it satisfies conditions i) and ii) of full RE states (i) being trivially satisfied by global optima). For the relative share we can consider ratio between the number of full RE states among global optima (fixed points) and the total number of global optima (fixed points) per configuration.

Since (full) account plays a prominent role in the definition of an RE state, the corresponding weight $alpha_A$ in heatmaps.

**Results** Observations:

- The relative share of full RE states among global optima always exceeds its counterpart for fixed points.
- The relative share of full RE states among global optima (fixed points) in linear models is roughly 2 times larger than in quadratic variants for all ensembles.
- The relative share of full RE states among global optima varies between 13.8% (QDS, ensemble 07) and 45.5% (LPS, ensemble 09)
- The relative share of full RE states among fixed points varies between 9.7% (QPS, ensebmle 08) and 30.1% (LDS, ensemble 09)
- There are not notable differences between pure and default systematicity in linear (quadratic) model variants.

- Across all model variants and ensembles, the weight for account has a positive impact on the mean relative share of full RE states among global optima.
- The heatmaps for fixed points differ from those for global optima in a peculiar manner. In contrast to global optima, the best values occur for non-extreme weights for account and faihtfulness in quadratic models. In addition to that, linear models also have high relative shares for low faihtfulness and high (but non-extreme) weights for systematicity.
- The heatmaps for the mean relative share of full RE states among global optima in linear model exhibit the typical "tipping line". Surprisingly and in contrast to the commitment consistency cases (see below), this does not hold for fixed points.
- Below the "tipping line" in linear models, all global optima are full RE state. This is not only an observation, but also a fact that can be proven analytically.
- The values in quadratic models tend to transition smoothly between different regions in heatmaps.

### Conclusion

Overall, the share of full RE states among global optima and fixed points is not overwhelming, but heatmaps reveal (plausible) combinations of weights for every model, where the relative share is acceptable (e.g $> 60\%$) Moreover, a low share of full RE states can be seen as a strength of a model, as it does not render everything into a full RE state.

## C) Commitment consistency cases

### Background

Consistency is commonly seen as a necessary condition of coherence, thus, if RE is taken to involve coherentist aspects, achieveing consistency is of utmost importance. This study considers the *dialectical consistency* (opposed to minimal consistency, i.e. the absence of flat contradictions, and the dialectical consistency of commitments plus a theory) of initial and endpoint commitments. The endpoint commitments stem either from a global optima or from a fixed point. The juxtaposition of initial and endpoint commitments allows for four cases, which are labelled as follows:

| | endpoint commitments consistent | endpoint commitment inconsistent |
|---|---|---|
| initial commitments consistent | good | very bad |
| initial commitments inconsistent | very good | bad |

Good case preserve or "transfer" consistency between intial and enpoint commitments. In very good cases, inconsistent initial commitments are revised for consistent endpoint commitments. Bad cases fail to eradicate initial inconsistencies and it is very bad if inconsistencies are introduced to initially consistent commitments.

### Method

During the generation of an ensemble we store the consistency status of intial commitments as well as the status for every commitment from global optima or fixed points for every configuration, and label them accordingly. Again, the number of global optima (fixed points) for a case type is put into relation with the total number of global optima (fixed points) for every configuration.

Good cases are not included in the consolidation. Surely, good cases are a desirable feature of RE, but the other cases seem to be more relevant/interesting for consolidating the models, because the strengths or weaknesses of RE are often discussed for such cases. Since commitment consistency cases are exhaustive, good cases take up the rest.

### Results

*Observations: Very good cases*

- The relative share of very good cases among global optima and fixed points varies between 10.3% (QDS, ensemble 08) and 22.7% (LDS, ensemble 07)
- Linear models have higher relative shares than quadratic variants in ensembles 08 and 09.
- The relative share of very good cases among global optima does not exceed its counterpart for fixed points (In ensemble 09, the converse holds.)
- linear model have a "tipping line" for very good cases among both global optima and fixed points
- very good cases occur only below the "tippling line" in linear models
- the mean relative share (and the standard deviation) below the tipping line in linear model are completely uniform.
- all model variants have regions in there heatmpas, where no very good cases occur at all, the maximal relative share of is 37% (ensebmle 07 and 08)/ 47% (ensemble 09) for both global and fixed points
- quadraatic models have smooth transitions
- in quadratic models, high weights for account and low weight for faithfulness benefit the relative share of very good cases among global optima and fixed points

*Observations: Bad cases*

- The relative share of bad cases among global optima varies between 14.8% (LDS, ensemble 08) and 33.8% (QDS, ensemble 09)
- quadratic models have a higher share of bad cases than linear variants in ensembles 08 and 09.
- The relative share of bad cases among global optima tend to exceed its counterpart for fixed points (In ensemble 09, the converse holds for linear models.)
- linear model have a "tipping line" for bad cases among both global optima and fixed points
- bad cases occur only above the "tippling line" in linear models
- the mean relative share (and the standard deviation) above the tipping line in linear model are completely uniform.
- all model variants have regions in there heatmpas, where no bad cases occur at all, the maximal relative share of is 37% (ensebmle 07 and 08)/ 47% (ensemble 09) for both global and fixed points
- quadraatic models have smooth transitions

- in quadratic models, high weights for faithfulness and low weight for account increase the relative share of bad cases among global optima and fixed points.

*Observations: Very bad cases*

- The relative share of very bad cases among global optima varies between 1.3% (LDS, ensemble 07) and 13.0% (QDS, ensemble 09)
- The relative share of very bad cases among fixed points varies between 0.4% (LPS, ensemble 07) and 8.7% (QDS, ensemble 08)
- quadratic models have a higher share of very bad cases than linear variants for both global optima and fixed points in all ensembles.
- The relative share of very bad cases among global optima exceed its counterpart for fixed points
- linear model have a "tipping line" for very bad cases among both global optima and fixed points
- very bad cases occur only above the "tippling line" in linear models
- all model variants have regions in there heatmpas, where no very bad cases occur at all
- quadraatic models have smooth transitions
- in quadratic models, low weights for account increase the relative share of very bad cases among global optima and fixed points.

**Conclusion**

Overall, linear models tend to perform better than quadratic variants for all cases, but every model variant has combinations of weights where very (bad) cases disapear and very good cases occur more often. Even very bad cases cannot serve as an exclusion criteria. They have a (very) small share and their manifestation may depend on additional input features (e.g. the size of initial commitmen or their minimal axiomatizsation).

## D) Extreme values for account and faitfulness

**Background** In this study we examine extreme values from measures (not the weights) of account and faithfulness.

$A(C, T) = 1$ means that the theory $T$ fully and exclusively accounts for the commitments $C$. Full and exclusive account is a condition for full RE states. Conversely, $A(C, T) = 0$ holds if a theory completely fails to account for commitments, that is, for every sentence in the commitments the closure theory does not contain this sentence, or contradicts the theory. $F(C|C_0) = 1$ holds iff the initial committments $C_0$ are a subset of the commitments $C$ (expansions of the initial commitments are not penalized). $F(C|C_0)$ attains the minimal value of 0, if every sentence of the initial commitments $C_0$ is missing in or contradicted by the commitments $C$.

Extreme values for systematicity are not included. $S(T) = 1$ holds if and only if $T$ is a singleton. $S(T) = 0$ is no serious option, because it is reserved for the empty theory $T = \emptyset$.

The values for account and faithfulness are calculated for every global optimum and fixed point from a configuration.

**Results**

Linear models reach maximal account and faithfulness values more frequently than quadratic variants in all ensembles for both global optima and fixed points

**Conclusion**

The frequency of minimal values for account and faithfulness in LDS is extremely small and they occur in one ensemble (07), only. Moreover, the extreme values for faithfulness and systematicity correlate with extreme, coresponding weights ($\alpha_F \leq 0.2$ and $\alpha_A \leq 0.2$).

The inclusion of non-extreme, but very low and very high values for account and faithfulness might provide a clearer picture of the situation.

The occurence of singleton theories in global optima and fixed points might make for an intersting exploration in the future.

## E) Consistency of theory and commitments

**Background**

A weak requirement of an RE state (and a consequence of full and exclusive account required for full RE) is that commitments and theory of a fixed point or a global optimum have to be consistent with each other, i.e. the union of commitments and theory has to be dialectically consistent.

**Method**

We store the consistency status of the union of commitments and theory for every fixed point and global optima resulting from a configuration during the creation of an ensemble.

**Results** *Observations:* - In ensemble 07 the relative share of consistent unions among both global optima and fixed points is roughly 75%. For the other ensembles, linear models perform slightly bettter than quadratic variants for global optima: 56% (QDS), 66% (QPS) vs. 79% (LDS, LPS) in ensemble 08, and 46% (QDS), 57% (QPS) vs. 61% (LDS), 64% (LPS) in ensemble 09. - there is no salient difference in the relative share of consistent unions for fixed points and global optima - the difference of relative shares of consistent commitments exceeds the one of consistent unions by roughly 5% (percentage points).

### Conclusion

The observed excess indicates, that in roughly 5% of cases for all model variants, the commitments of fixed points or global optima are consistent by themselves, but inconsistent with the theory. It may be interesting to explore, whether there are other aspects of configurations (e.g. low weight for account) that provoke such cases.

## F) Process Lenght

### Background

The length of an RE process is defined as the number of adjustment steps that bring about a change in the commitments or the theory. In contrast to the implemented stopping rule (check after even-numbered steps, and inclusion of steps without change), this definition is closer to a intuitive understanding of how long an RE process is.

### Method

### Results

A noteworthy finding in all ensembles is the short length of RE processes, which is even more accentuated for linear models. In ensemble 06 (including the size of jumps in individual steps), 78.9% of processes have lenght 2, and 99.6% have length $\leq 3$.

### Conclusion

In view of the semi-global approach to adjustment steps (i.e. the consideration of all candidates for theories or commitments in contrast to genuine piece-meal adjustment of single elements) short processes are expectable and do not speak in favour or against a model variant.

## G) Principles

### Background

The new generation of random arguments allows to ensure that the dialectical structures includes principles, i.e. sentences that only occur as premises in arguments. Consequently, principles have potential to account for other sentences (maybe together wit auxiliary premises). The question is whether fixed points or global optima tend to include principles in their theories.

### Method

*Definition*: A sentence is a principle of multiplicity $n$ in a dialectical structure iff (i) it occurs in exactly $n$ arguments as a premise and (ii) it or its negation does not occur as a conclusion in any argument.

We generated ensemble 06 with at least two principles per dialectical structure (150 structures). Upon initialization, the principles and their multiplicity are saved as a list of tuples of the form (principle, multiplicity) in the dataframe. The study only included principles with multiplicity ` $\geq$` 2 and addressed the following questions: - Comparison with a chance model for the expected number of principles occurring in if theories were selected randomly. - Are principles in initial commitments preserved/transferred to fixed point theories?

**Results**

*Observations* - the number of principles with multiplicity $\geq 2$ is normally distributed around 2.

*Chance model* - in most cases the mean number of principles in fixed point theories is higher than the expected number by chance

*Preservation of principles* - in most of processes the amount of principles does not change from initial commitments to fixed point theory. - QDS performs better with respect to the mean difference of principles in initial commitments and theory (QDS: 0.0) than the other variants (QPS, LDS, LPS: -0.18) - the difference of principles in intial commitments and theory depends on the number of principles in the initial commitments. Only for initial commitments without principles the difference is positive. Again, QDS performs best.

**Conclusion**

The results are interesting but not decisive or well understood for the consolidation of the model.There are many open questions concerning the current implementation of principles:

- Is the definition of principles reasonble in view of alternatives?
- Is it possible to define other important and related concepts syntactically (e.g. additional or background assumptions, evidence etc.)?
- Is the chance model too benevolent?
- Should we consider only those principles, which are not elements of the initial commitments?
- Should the chance model consider the distribution of principles in the set of all consistent and complete positions?

In contrast to later ensembles, ensemble 06 does not completely track all fixed points and global optima for a configuration. It may be interesting to explore principles in a newly generated ensemble, which allows to study all fixed points as well as global optima.

## H) Replication of published Results

**Background**

The published paper includes a study based on a Mathematica implementation of Gregor Betz ((GitHub reposistory)[https://github.com/debatelab/remoma]).

**Method** The studies of the pupblished paper include two ensembles.

**Results**

The Python implementation of the published model (QuadraticDefaultSystematicityRE) is able to replicate the results from the Mathematica implementation from the published paper.

**Conclusion**

## Literature

Beisbart, Claus, Gregor Betz, and Georg Brun. 2021. "Making Reflective Equlibrium Precise: A Formal Model." *Ergo* 8 (0). https://doi.org/10.3998/ergo.1152.

Betz, Gregor. 2010. *Theorie dialektischer Strukturen*. Frankfurt am Main: Klostermann.

———. 2013. *Debate Dynamics: How Controversy Improves Our Beliefs*. Synthese Library. Dordrecht: Springer Netherlands.