Simple models of deliberated preference

Olivier Cailloux

27th November, 2017

1. Motivation

We define the deliberated preference of an individual. We are interested in a context of a choice of a subset of alternatives from a set of possible alternatives given a priori. We define a model of the DP and give conditions and a validation procedure that ensure that what the model claims is in the DP is indeed. We want in particular to allow for the model to be applicable to various individuals that start with different knowledge: the model may use different counter-arguments to convince different individuals.

More in detail. A full-blown proof may be so long that nobody would read it. (That's why mathematicians do not write their proofs in set-theoretic bare notation.) Hence, shortcuts are taken, which rely on the individual knowing stuff already. More generally, assume you want to convince someone of fact F with a text (we call this an argument), in our case, we want to convince a person that such alternative from the set is a good choice for her. The argument may assume the reader knows already why such or such possible counter-arguments are invalid, and the argument will thus not delve into the details of some points. When shown in interactive mode (our context here), an individual may however question some points (which we call a counter-argument). The argumenter should then be able to provide an answer to the query (a counter-counter-argument).

This article aims at defining such a procedure and conditions of its validity.

2. Decision situation

Here is the definition of a decision situation.

- All alternatives $\mathscr{A}$.
- Topic $T^* = \left\{ t_a, a \in \mathscr{A} \right\} \cup \left\{ t_{\neg a}, a \in \mathscr{A} \right\}$. Denoted simply a and $\neg a$. We define $\neg t$, with $t = t_a$, as equal to $t_{\neg a}$ and $\neg t$, with $t = t_{\neg a}$, as equal to t_a.
- All possible arguments: S^*. It contains at least $\emptyset$, the empty argument. It contains all the arguments and more. As an example, S^* could be the set of all strings (assuming all arguments can be transmitted textually).

Define f_t from any unordered pair of arguments (s, s') into $\mathscr{P}(t\text{-sure}, \neg t\text{-sure}, b))$ (where b means both possible). We present both arguments to i and ask which proposition i considers valid in her current state of mind, thus using both arguments and possibly other arguments she has in mind. The set designates all cumulated answers over time, thus allowing i to be unstable (because she has learnt new things in between two questioning, or just because of time passing, or for any known or unknown reason). We assume b is observable as well. We can also use $\emptyset$ as one or both of the arguments.

We define two binary relations t^+ (always supports t-poss) and t^- (always supports $\neg t$-sure) on S^*. We define $(s, s') \in t^+ \Leftrightarrow (s, s') \overset{\text{poss}}{\rightsquigarrow}_\forall t \Leftrightarrow s' \ntriangleright^t_\forall s \Leftrightarrow \rightsquigarrow(s, s') \subseteq \{t\text{-sure}, b\}$, and $(s, s') \in t^- \Leftrightarrow (s, s') \overset{\text{sure}}{\rightsquigarrow}_\forall \neg t \Leftrightarrow s \ntriangleright^{\neg t\text{-sure}}_\forall s' \Leftrightarrow \rightsquigarrow(s, s') = \{\neg t\text{-sure}\}$. Hence, $(s, s') \notin t^+ \Leftrightarrow (s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t \Leftrightarrow s' \triangleright^t_\exists s \Leftrightarrow \neg(s' \ntriangleright^t_\forall s) \Leftrightarrow \rightsquigarrow(s, s') \supseteq \{\neg t\text{-sure}\}$, and $(s, s') \notin t^- \Leftrightarrow (s, s') \overset{\text{poss}}{\rightsquigarrow}_\exists t \Leftrightarrow s \triangleright^{\neg t\text{-sure}}_\exists s' \Leftrightarrow \neg(s \ntriangleright^{\neg t\text{-sure}}_\forall s') \Leftrightarrow \rightsquigarrow(s, s') \neq \{\neg t\text{-sure}\}$.

When t is fixed, or if it can be deduced from the output set, we write $(s, s') \rightsquigarrow O$ or $\rightsquigarrow(s, s') = O$ instead of $f_t(s, s') = O$.

Actually we will use only the $\overset{\text{poss}}{\rightsquigarrow}_\forall$ and $\overset{\text{sure}}{\rightsquigarrow}_\forall$ ternary relations from now on, together with the property (A1) that $\overset{\text{poss}}{\rightsquigarrow}_\forall \subseteq \overset{\text{poss}}{\rightsquigarrow}_\exists$ [1] and (A2) that $\overset{\text{sure}}{\rightsquigarrow}_\forall \subseteq \overset{\text{poss}}{\rightsquigarrow}_\forall$ [2] and the fact

[1] Equivalently: $\overset{\text{sure}}{\rightsquigarrow}_\forall \subseteq \overset{\text{sure}}{\rightsquigarrow}_\exists$, $(s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t \vee (s, s') \overset{\text{poss}}{\rightsquigarrow}_\exists t$, $\overset{\text{poss}}{\rightsquigarrow}_\forall t \Rightarrow \neg(\overset{\text{sure}}{\rightsquigarrow}_\forall \neg t)$, $\overset{\text{sure}}{\rightsquigarrow}_\forall \neg t \Rightarrow \neg(\overset{\text{poss}}{\rightsquigarrow}_\forall t)$.

[2] Equivalently: $\overset{\text{sure}}{\rightsquigarrow}_\exists \subseteq \overset{\text{poss}}{\rightsquigarrow}_\exists$.

that the relations t^+ and t^- are symmetric. [3]

We say (s, s') is t-unstable iff $(s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists t \wedge (s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t$.

s is

- t-sure-dec $\Leftrightarrow \forall s' : (s', s) \overset{\text{sure}}{\rightsquigarrow}_\forall t$
- t-poss-dec $\Leftrightarrow \forall s' : (s', s) \overset{\text{poss}}{\rightsquigarrow}_\forall t$

We can now define $T_i, T_i^{\text{sure}} \subseteq T^*$:

- $t \in T_i \Leftrightarrow \exists s$ t-poss-dec. [4]
- $t \in T_i^{\text{sure}} \Leftrightarrow \exists s$ t-sure-dec. [5]

We say that severe t-unstability happens iff $\forall s, [\exists s^+ \mid (s, s^+) \overset{\text{sure}}{\rightsquigarrow}_\exists t] \wedge [\exists s^- \mid (s, s^-) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t]$. [6] [7]

Theorem 1. *If $t \notin T_i \wedge \neg t \notin T_i$, severe t-unstability happens.*

The proof follows from footnote 4.

Theorem 2. $t \in T_i^{sure} \Rightarrow \neg t \notin T_i$.

This is because s^+ is t-sure-dec, thus given any s^-, $(s^+, s^-) \overset{\text{sure}}{\rightsquigarrow}_\forall t$ hence $(s^+, s^-) \overset{\text{sure}}{\rightsquigarrow}_\exists t$ (using A1).

Theorem 3. $t \in T_i^{sure} \Rightarrow t \in T_i$.

The proof follows from (A2).

Definition 1 (Clear-cut)**.** *A situation is t-clear-cut iff $t \notin T_i \Rightarrow \neg t \in T_i^{sure}$.* [8] [9]

[3] Unclear where we use the fact that the relations t^+ and t^- are symmetric.

[4] $t \notin T_i \Leftrightarrow \forall s, \exists s' \mid (s', s) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t$.

[5] $\neg t \notin T_i^{\text{sure}} \Leftrightarrow \forall s, \exists s' \mid (s', s) \overset{\text{poss}}{\rightsquigarrow}_\exists t$.

[6] Severe t-unstability happens iff severe $\neg t$-unstability happens.

[7] Here is an example of severe t-unstability. $\forall (s, s'), (s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t, (s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists t$.

[8] Here is an example of non t-clear-cut situation with no severe t-unstability. $\forall (s, s'), (s, s') \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t, (s, s') \overset{\text{poss}}{\rightsquigarrow}_\exists t, (s, s') \overset{\text{poss}}{\rightsquigarrow}_\forall \neg t$.

[9] Something like Axiom JU should be sufficient for CC. JU: If (s, s') is t-unstable, then $\exists s_2 \in S^*$ such that for $(s_1, s_0) = (s, s')$ or for $(s_1, s_0) = (s', s)$, s_2 attacks decisively s_1, where we mean by this that s_2 combined with s_0 is decisive. Means that if not CC, then not JU. Hence, infinite discussion happens, or irreducible unstability. Or, the memory buffer of i is too short to hold adequate reinstated arguments, in which case our approach based on decisive arguments is not adequate.

2.1. Possible outcomes

We know $t \in T_i^{\text{sure}} \Rightarrow \neg t \notin T_i$ and $t \in T_i^{\text{sure}} \Rightarrow t \in T_i$.

Here are the remaining possibilities, considering the four pairs (t, T_i), (t, T_i^{sure}), $(\neg t, T_i)$, $(\neg t, T_i^{\text{sure}})$. (A pair (t, T) not mentioned means $t \notin T$.)

All poss	$t \in T_i$, $\neg t \in T_i$
t **sure**	$t \in T_i$, $t \in T_i^{\text{sure}}$
$\neg t$ **sure**	(symmetric)
Unstability, t **poss**	$t \in T_i$ (oddity: $\neg t \notin T_i$ should imply $t \in T_i^{\text{sure}}$)
Unstability, $\neg t$ **poss**	(symmetric)
Strong unstability	(both oddities above)

3. Models

A model is a triple $(\rhd_\eta, \rightsquigarrow_\eta, +)$ defined as follows and satisfying the constraints as indicated here.

$\rhd_\eta$ a binary relation over S^*. $\rightsquigarrow_\eta \subseteq S^* \times T^*$. Define $S_\eta \subseteq S^*$ as the set of arguments used in $\rhd_\eta \cup \rightsquigarrow_\eta$. Let $+$ be defined over arguments used in the model: $s_3 + s_1 = s'$ for some $s' \in S_\eta$, for any $s_3, s_1 \in S_\eta$.

Requirements. The maximum length of a path in $\rhd_\eta$ is finite (which implies that $\rhd_\eta$ is acyclic). $s_3 \rhd_\eta s_2 \rhd_\eta s_1 \Rightarrow s_2 \rhd_\eta s_3 + s_1$. [10]

Notation. Let $\rightsquigarrow_\eta^{-1}(T^*) \subseteq S_\eta$ denote the subset of arguments supporting propositions. $\rhd_\eta(s_2)$: arguments that s_2 attacks, $s_1 \in \rhd_\eta(s_2) \Leftrightarrow s_2 \rhd_\eta s_1$. We write $S \rhd_\eta s$ to mean that $\forall s' \in S : s' \rhd_\eta s$, and similarly for other binary relations.

Given a decision situation, define $\succ_\exists^t, \not\succ_\exists^{\neg t} \subseteq \rhd_\eta$ as follows.

Given $s_3 \rhd_\eta s_2, t \in T^*$: $s_3 \succ_\exists^t s_2$ iff $[\exists s_1 \in \rhd_\eta(s_2) \mid (s_2 \succ_\exists^t s_1 \wedge s_2 \not\succ_\exists^t s_3 + s_1)] \vee [s_2 \rightsquigarrow t \wedge (s_3, s_2) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t]$. [11]

Given $s_3 \rhd_\eta s_2, t \in T^*$: $s_3 \not\succ_\exists^t s_2$ iff $[\exists s_1 \in \rhd_\eta(s_2) \mid (s_2 \succ_\exists^t s_1 \wedge s_2 \succ_\exists^t s_3 + s_1)] \vee [s_2 \rightsquigarrow_\eta t \wedge (s_3, s_2) \overset{\text{poss}}{\rightsquigarrow}_\exists t]$.

This is well defined thanks to the following construction. Given a model.

- Define a root as an argument that $\rhd_\eta$-attacks nobody.

[10] Necessary for definition of $s_3 \succ_\exists s_2$.

[11] Or $s_2 \rightsquigarrow t \wedge \neg t \notin \rightsquigarrow_\eta(S^*) \wedge (s_3, s_2) \overset{\text{poss}}{\rightsquigarrow}_\exists \neg t$.

- Associate to each $s \in S_\eta$ its depth $d(s)$, the distance to the farthest root: $d(s) = 0$ iff s is a root and $d(s) = k+1$ iff $s \rhd_\eta$-attacks some s' of depth k and $\rhd_\eta$-attacks no s' of depth superior to k (thus $d(s) = 1$ iff s attacks some and only root nodes; $d(s) = 2$ iff s attacks some node of depth 1 and only nodes of depth 1 or 0; and so on).

- Observe that if $s_3 \rhd_\eta s_2$ and $d(s_3) = k + 1$, then $d(s_2) \leq k$.

- Define $\succ^t_{\exists(0)} = \not\succ^t_{\exists(0)} = \emptyset$.

- Define $\succ^t_{\exists(k+1)} \subseteq \rhd_\eta$, $k \geq 0$, as $s_3 \succ^t_{\exists(k+1)} s_2$ iff, first, $d(s_3) = k + 1$, and second, $[\exists s_1 \in \rhd_\eta(s_2) \mid (s_2 \succ^t_{\exists(d(s_2))} s_1 \land s_2 \not\succ^t_{\exists(d(s_2))} s_3 + s_1)] \lor [s_2 \rightsquigarrow t \land (s_3, s_2) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t]$. Proceed similarly for $\not\succ^t_{\exists(k+1)}$.

- Define $\succ^t_\exists = \bigcup_{k\geq 0} \succ^t_{\exists(k)}$, and similarly for $\not\succ^t_\exists$.

Check: Given $s_4 \rhd_\eta s_3, s_3 \notin \rightsquigarrow^{-1}_\eta(T^*)$, with $s_2 \in \rhd_\eta(s_3) \Rightarrow$ (s_2 and $s_4 + s_2$ $\rhd_\eta$-attack only root nodes and do not support any proposition): $s_4 \succ^t_\exists s_3$ iff

- $\exists s_2 \in \rhd_\eta(s_3) \mid [(s_3 \succ^t_\exists s_2) \land (s_3 \not\succ^t_\exists s_4 + s_2)]$ iff

- $\exists s_2 \in \rhd_\eta(s_3) \mid [(\exists s_1 \in \rhd_\eta(s_2) \mid s_2 \succ^t_\exists s_1 \land s_2 \not\succ^t_\exists s_3 + s_1) \land (\exists s_1 \in \rhd_\eta(s_4 + s_2) \mid s_4 + s_2 \succ^t_\exists s_1 \land s_4 + s_2 \succ^t_\exists s_3 + s_1)]$ iff

- $\exists s_2 \in \rhd_\eta(s_3) \mid [(\exists s_1 \in \rhd_\eta(s_2) \mid s_1 \rightsquigarrow t \land (s_2, s_1) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t \land s_3 + s_1 \rightsquigarrow t \land (s_3 + s_1, s_2) \overset{\text{poss}}{\rightsquigarrow}_\exists t \land (\exists s_1 \in \rhd_\eta(s_4 + s_2) \mid s_1 \rightsquigarrow_\eta t \land (s_4 + s_2, s_1) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t \land s_3 + s_1 \rightsquigarrow t \land (s_4 + s_2, s_3 + s_1) \overset{\text{sure}}{\rightsquigarrow}_\exists \neg t)]$.

Given $s_3 \in S_\eta, s_2 \in S_\eta$, with $\exists s_1 \in \rhd_\eta(s_2) \mid s_2 \succ^t_\exists s_1$, we have: $s_3 \succ^t_\exists s_2 \lor s_3 \not\succ^t_\exists s_2$.

Define $s_2 \succ_\exists s_1 \Leftrightarrow \exists t \in T^* \mid s_2 \succ^t_\exists s_1$.

Hence, given $s_3 \in S_\eta, s_2 \in S_\eta, s_2 \notin \rightsquigarrow^{-1}_\eta(T^*)$: $\neg(s_3 \not\succ_\exists s_2)$ iff $\forall s_1 \in \rhd_\eta(s_2) \cap \succ_\exists(s_2) : \neg(s_2 \succ_\exists s_3 + s_1)$.

4. Conditions

All these conditions assume that a decision situation $(\mathscr{A}, S^*, \{\rhd^t_\exists\}, \{\rhd^{t,\text{sure}}_\exists\})$ and a model $\eta = (\rhd_\eta, \rightsquigarrow_\eta, +)$ are given.

Define $S_{\text{decisive}} = S_\eta \setminus \text{im}(\succ_\exists)$ the decisive arguments according to $\succ_\exists$, or $\succ_\exists$-decisive arguments for short: $s \in S_{\text{decisive}} \Leftrightarrow \succ^{-1}_\exists(s) = \emptyset$.

Definition 2 (Reinstatement)**.** *Given $s_3 \succ_\exists s_2 \succ_\exists s_1, s_3 \in S_{\text{decisive}}$: $\rhd_\eta(s_1) \subseteq \rhd_\eta(s_3+s_1) \land \rhd^{-1}_\eta(s_3+s_1) \subseteq \rhd^{-1}_\eta(s_1) \setminus \rhd_\eta(s_3)$.* [12] [13] [14]

Definition 3 (Justifiable unstability)**.** $\forall s_2 \rhd_\eta s_1 \mid s_2 \succ_\exists s_1, s_2 \not\succ_\exists s_1 : \exists s_3 \rhd_\eta s_2 \mid s_3 \succ_\exists s_2$.

Definition 4 (Finite defense)**.** *If $\succ^{-1}_\exists(s) \subseteq \succ_\exists(S_{\text{decisive}})$, then $\exists S \subseteq S_{\text{decisive}}, |S| \leq j \mid \rhd^{-1}_\eta(s) \subseteq \succ_\exists(S)$.* [15]

Thus, if the attackers of s are attacked by decisive arguments, then j defenders are enough to defend s.

Define R, the reinstates relation, as follows: $s_3 R s_1$ iff $s_3 \succ_\exists s_2 \succ_\exists s_1$ (for some s_2), $s_3 \in S_{\text{decisive}}$. Define S_γ as the transitive closure of $\rightsquigarrow^{-1}_\eta(T^*)$ under R.

Definition 5 (Covering)**.** $\forall s \in S_\gamma, s' \in S^* : s' \rhd_\exists s \Rightarrow s' \rhd_\eta s$. [16]

Definition 6 (Observable validity)**.** $\forall s_2 \rhd_\eta s_1 \rightsquigarrow_\eta t : \neg(s_2 \succ^t_\exists s_1) \lor \exists s_3 \rhd_\eta s_2 \mid s_3 \succ^t_\exists s_2$. *Furthermore, if $\neg(S_\eta \rightsquigarrow_\eta t), \forall s_1 \rightsquigarrow_\eta \neg t, s \in S^* : s_1 \rhd^t_\exists s$.* [17]

[12] TODO the condition must be $\succ^{-1}_\exists(s_3 + s_1) \subseteq \rhd^{-1}_\eta(s_1) \setminus \rhd_\eta(s_3)$ to allow $s_5 \rhd_\eta s_4 \rhd_\eta s_3 \rhd_\eta s_2 \rhd_\eta s_1$ and $s_5 \rhd_\eta s_4 \rhd_\eta s_3 + s_1$, considering that possibly s_3 is $\succ_\exists$-decisive. This should not invalidate the conditions, but it does currently. But it's not a problem: the model would actually not be built this way. In this scenario the argument $s_3 + s_1$ is useful only in case s_3 is decisive, thus $s_4 \rhd_\eta s_3 + s_1$ must not be planned. Rather $s_5 + s_3$ decisive, then $(s_5 + s_3) + s_1$. Alternatively, also $s_4 \rhd_\eta s_1$ and then no problem as well.

[13] The stronger condition mandating $\rhd^{-1}_\eta(s_3 + s_1) \subseteq \succ^{-1}_\exists(s_1) \setminus \rhd_\eta(s_3)$ would be more difficult to check: when some $s_2 \succ_\exists s_3 + s_1$, we'd need to check not only that $s_2 \rhd_\eta s_1$ but also $s_2 \succ_\exists s_1$.

[14] We do not mandate that $s_3 + s_1 \rightsquigarrow t$, so that the model can afford not resisting to the counter-attacks to $s_3 + s_1$ (resistance to c-a to s_1 suffice). We need: Obs applies to restricted supports (one per prop decided by model); Covering applies to extended supports (restricted supports plus those obtained by reinstatement). Replacement-1 applies to all and requires attack at least as large; Replacement-2 applies to restricted supports and requires no new $\rhd_\exists$-attacks.

[15] To satisfy Finite defense, in presence of the other conditions, suffice to limit the width of the model (TODO check). But it may be interesting to not limit it and declare that the model has specific replies to any counter-argument, but promises to use only a few rebuttals and that afterwards, the dm will stop using those kind of arguments (but we don't know in advance which ones will be chosen).

[16] Specify $\rhd_\exists$.

[17] If the model claims $\neg t \in T^{\text{sure}}_i$, this requires clear-cut (for that prop), so we must mandate it (hopefully A3 or an equivalent such as Justifiable unstability fits). Thus we only need to prove $t \notin T_i$, for which $s_1 \rhd^t_\exists s$ suffices.

5. Theorem

Theorem 4 (Validity). *Given a decision situation and a model η, if all our conditions are satisfied, $\rightsquigarrow_\eta(S_\eta) \subseteq T_i$. Furthermore, if $\neg t \in \rightsquigarrow_\eta(S_\eta) \land t \notin \rightsquigarrow_\eta^{-1}(S_\eta)$, $\neg t \in T_i^{sure}$.*

Proof. s is defended iff its $\succ_\exists$-attackers are $\succ_\exists$-attacked by $\succ_\exists$-decisive arguments.

First, we want to prove that s_1 defended implies s_1 replaceable by some $\succ_\exists$-decisive s, and if $s_1 \in S_\gamma$, then its replacer s is in S_γ as well.

By hypothesis, $\succ_\exists^{-1}(s_1) \subseteq \succ_\exists(S_{\text{decisive}})$. Thus, $\exists S \subseteq S_{\text{decisive}} \mid \rhd_\eta^{-1}(s_1) \subseteq \succ_\exists(S)$, S finite [Finite defense].

Pick any $s_{3,1} \in S$ such that $s_{3,1} \succ_\exists s_2 \succ_\exists s_1$ (if there's none, $s_1 \in S_{\text{decisive}}$ and we're done). $s_{3,1} + s_1$ replaces s_1, and $\rhd_\eta^{-1}(s_{3,1} + s_1) \subseteq \rhd_\eta^{-1}(s_1) \setminus \rhd_\eta(s_{3,1})$ [Reinstatement]. Hence, $\rhd_\eta^{-1}(s_{3,1} + s_1) \subseteq \succ_\exists(S) \setminus \rhd_\eta(s_{3,1})$. Iterate by picking any $s_{3,2} \in S$ such that $s_{3,2} \succ_\exists s_2 \succ_\exists s_{3,1} + s_1$ (if there's none, $s_{3,1} + s_1 \in S_{\text{decisive}}$ and we're done) and obtaining $s_{3,2} + (s_{3,1} + s_1)$ replacing $s_{3,1} + s_1$ (hence, replacing s_1) with $\rhd_\eta^{-1}(s_{3,2} + (s_{3,1} + s_1)) \subseteq \rhd_\eta^{-1}(s_{3,1}+s_1) \setminus \rhd_\eta(s_{3,2})$. Hence, $\rhd_\eta^{-1}(s_{3,2}+(s_{3,1}+s_1)) \subseteq \succ_\exists(S) \setminus \rhd_\eta(s_{3,1}) \setminus \rhd_\eta(s_{3,2})$. Iterating in such a way over the finite set S will finally yield an element that is $\succ_\exists$-decisive. The last point, $s_1 \in S_\gamma \Rightarrow s \in S_\gamma$, follows from the definition of S_γ.

Second, we want to prove that if s_1 not defended and has no decisive $\succ_\exists$-attackers (meaning that $\succ_\exists^{-1}(s_1) \subseteq \overline{S_{\text{decisive}}}$), then s_1 is $\succ_\exists$-attacked by some s_2 that is not defended and has no decisive $\succ_\exists$-attacker.

Consider s_1 not defended and having no decisive $\succ_\exists$-attackers. Because s_1 is not defended, by definition, it is $\succ_\exists$-attacked by some s_2 that has no decisive $\succ_\exists$-attacker. Because s_1 has no decisive attacker, s_2 is not decisive. If s_2 was defended, by the first part of this proof, it would be replaceable by a decisive argument, and s_1 would have a decisive attacker. Thus, s_2 is not defended.

Third, consider an argument $s_1 \in \rightsquigarrow_\eta^{-1}(T^*)$. It has no decisive $\rhd_\exists$-attacker: as $s_1 \in S_\gamma$, any $\rhd_\exists$-attack is a $\succ_\exists$-attack [Covering], and s_1 has no decisive $\succ_\exists$-attacker [Obs val]. Also, s_1 is defended: assume it is not, then by our second point in this proof some $s_2 \succ_\exists s_1$, with s_2 not defended and with no decisive $\succ_\exists$-attacker, and iterating and using finiteness of $\succ_\exists$ leads to a contradiction. Hence, by our first point in this proof, s_1 is replaceable by some $\succ_\exists$-decisive $s \in S_\gamma$. As $s \in S_\gamma$, any $\rhd_\exists$-attack is a $\succ_\exists$-attack [Covering], thus s is $\rhd_\exists$-decisive. □

A. Todo

- If i does not consider s as supporting t, it also works: if t is not weakly acceptable by default, then any s' is considered by i as a better argument than s in favor of certain $\neg t$, and so on. In fact, whether $\varnothing \rhd_\exists^t \varnothing$ determines whether t is weakly supported by default.

- I should define $s'(\Box\rhd_\exists^t)s$ as an observable: "Assuming s' would survive, do you consider s' as leading to certainty of $\neg t$, even when considering s?". It distinguishes our knowledge and the truth: $s'(\Box\rhd_\exists^t)s \Rightarrow s' \rhd_\exists^t s$, thus, implies $\neg(s' \not\rhd_\forall^t s)$. But out of $\neg(s'(\Box\rhd_\exists^t)s)$, nothing.

- Partition (objectively) S^* (or $S^* \times T^*$) into arguments in favor of t, sure, $\neg t$, sure, and similarly for possible. Use only one rel $\rhd_\exists$, defined on contradictory arguments only, instead of $\rhd_\exists^{t,\text{sure}}$ and others. Define $s' \rhd_\exists^t s$ equals no when $\neg(s' \rightsquigarrow \neg t, \text{sure})$, equals $\rhd_\exists$ for adequate arguments, and equals yes when $\neg(s \rightsquigarrow t, \text{possible})$ and $s' \rightsquigarrow \neg t$, sure, with probably some complications needed for the argument $\varnothing$ (and related default attitude towards t).

Questions: Q1. Relationship with $s \rhd_\exists^t s$?

We want to exclude: s supports p perhaps, attacked by $s2$ (supporting $\neg p$ sure), but then $s2$ is attacked by s. Exclude $s' \rhd_\exists^t s$ and $s \rhd_\exists^{\neg t,\text{sure}} s'$. Require to assume that this situation implies another argument s_3 "attacking" s', thus, such that $s_3 + s$ is no more attacked by s_2.

A.1. Attack rel

Given t, define a (symmetric) strong attack rel, Q_t, as sQ_ts' iff $(s, s') \overset{\text{sure}}{\rightsquigarrow}_\forall \neg t$.

Define a (symmetric) weak attack rel, R_t, as sR_ts' iff $(s, s') \overset{\text{poss}}{\rightsquigarrow}_\forall \neg t$.

Given t, we want to check the arguments that defend it. We want to build an asymmetric attack relation. $s_1 M s_0$, and not the converse, iff s_1 attacks s_0 and, would s_1 be attacked, s_0 reinstated would defend t. s_1 should be able to support a sub-claim t', and then s_2 attack s_1 on that sub-claim precisely. s_2 decisive on this sub-claim and $s_0 + s_2$ decisive implies s_1 attacks s_0 and not conversely.

Or simply. s_1 decisive for $\neg t$ means s_1 attacks s_0 and not conversely.

Then define i's DJ in terms of argumentation semantics.

B. To think

Propositions weakly self-supported $T \subseteq T^*$: weakly accepted if no arg is given. Examples: m = "eat miam"; $\neg b$ = "beurk is to exclude"; or, in a problem where there's no particularly good aliments, both a = "eat this" and $\neg a$.

When given (s, t), i may say: s does not survive; or: assuming s survives, then s supports t, or, assuming s survives, then s does not support t anyway.

When given s' against s, i may say: s' does not survive, or: assuming s' survives, then s' supports $\neg t$, …

Given $(s_2, t), (s_1, \neg t) \in D$, define $\neg(s_2 \succ^{\text{neg}}_{\exists\neg t} s_1)$ iff for some $(s, t) \in D$, where $s_1 \rhd^t_\exists s$: $s_1 \rhd^t_\exists s + s_2$. Equivalently: $s_2 \succ^{\text{neg}}_{\exists\neg t} s_1$ iff for all s, where $s_1 \rhd^t_\exists s$: $\neg(s_1 \rhd^t_\exists s + s_2)$. (This does not seem right: if given s_3 attacking s_2, and not given s_4 which would convincingly rebut s_3, then temporarily it may hold again that $s_1 \rhd^t_\exists s + s2$ (in the sense that $s_1 + s_3 \rhd^t_\exists s + s_2$).)

$s_2 \succ_{\exists\neg t} s_1$ can perhaps be queried directly by asking (in the context of some $s_1 \rhd^t_\exists s$): "assume s_2 survives, then does s_2 counter s_1?" (In the sense that s_2 is sufficiently convincing that t holds perhaps, to cancel the argument s_1 according to which $\neg t$ surely holds.)

C. Certainties

Looking for certainties. Those propositions that are in the reflexive preferences in a demanding sense: there is a strong enough reason to prefer it than its contrary.

- $s' \succ_\exists s$: weak attack; s' renders s invalid (can't be used to say that t holds for sure) (assuming s' survives)
- Propositions strongly self-supported: strongly accepted if no arg is given. Examples: m = "eat miam"; $\neg b$ = "beurk is to exclude". We might have neither c nor $\neg c$ in that set.

Definition 7 (Sure acceptance)**.** *Define a situation* $(\mathscr{A}, S^*, \{\rhd^t_\exists\})$*. A proposition* $t \in T^*$ *is accepted as sure iff* $\exists s' \in S^* \mid \forall s \in S^* : s \not\rhd^{t,sure}_\forall s'$.

Assume we use rather: if p is not sure, then $\neg p$ is weakly accepted (by def). Then we have never problems of inconsistency! But we could be in a situation where p is not accepted as sure but nobody can tell why because it is fundamentally unstable (sometimes p being accepted, sometimes not).

D. Example about model instanciation

The general conditions are Reinstatement, Justifiable unstability, Finite defense and Covering. A general model is a model that claims it satisfies the general conditions.

TODO give up general models. In this example, s_1 would need to be planned as attacking sometimes s_2. Better consider an instanciation mechanism. An instantiated model is particular, and can be tested (especially against another one).

Example 1. $s_3 \rhd_\eta s_2 \rhd_\eta s_1 \rightsquigarrow_\eta t$, $s_2 \rightsquigarrow_\eta \neg t$; $s_3 + s_1 \rightsquigarrow_\eta t$. △

This model is compatible (meaning that it satisfies the general conditions) with the following decision situations. We describe $\rhd_\exists$ fully (no attack iff not mentioned).

- Sure of t: $s_1 \rightsquigarrow_\eta t$; $s_3 \rhd_\exists s_2$ (the rest is implied, for example $\forall s_4 \in S^* : s_1 \rhd^{\neg t}_\exists s_4$ because of covering).
- Sure of t with reinstatement: $s_3 \rhd^t_\exists s_2 \rhd^t_\exists s_1 \rightsquigarrow_\eta t$; $s_1 + s_3 \rightsquigarrow_\eta t$; $s_3 \rhd^{\neg t}_\exists s_2$
- Sure of $\neg t$: $s_2 \rightsquigarrow_\eta \neg t$; $s_2 \rhd_\exists s_1$
- Both: $\neg(s_2 \rhd_\exists s_1), s_1 \rightsquigarrow_\eta t, \neg(s_3 \rhd_\exists s_2), s_2 \rightsquigarrow_\eta \neg t$

This situation falsifies the model. $s_4 \rhd_\exists s_1$, s_4 not attacked.

E. Model certainties

Assume we define $s_1 \succ^{\neg t,\text{sure}}_\exists s_2 \Leftrightarrow s_2 \not\succ^t_\exists s_1$. Then, indeed, given $s_1 \rightsquigarrow_\eta t$, $s_2 \succ^{\neg t,\text{sure}}_\exists s_1 \Leftrightarrow s_2 \rhd^{\neg t,\text{sure}}_\exists s_1$. But it gives the wrong conclusion. For $s_3 \rhd_\eta s_2 \rhd_\eta s_1 \rightsquigarrow_\eta t$: $s_3 \succ^{t,\text{sure}}_\exists s_2$ iff $s_2 \not\succ^{\neg t}_\exists s_3$ iff $\exists s_1 \in \rhd_\eta(s_3) \mid s_3 \succ^{\neg t}_\exists s_1 \wedge s_3 \succ^{\neg t}_\exists s_2 + s_1$.

Given $s_1 \rightsquigarrow_\eta t$, define $s_2 \succ^{t,\text{sure}}_\exists s_1$ iff $s_2 \rhd^{t,\text{sure}}_\exists s_1$.

Given $s_1 \rightsquigarrow_\eta t$, define $s_2 \not\succ^{t,\text{sure}}_\exists s_1$ iff $s_1 \rhd^{\neg t}_\exists s_2$.

Given $s_3 \in S_\eta, s_2 \in S_\eta, s_2 \notin \rightsquigarrow^{-1}_\eta(T^*), t \in T$: $s_3 \succ^{t,\text{sure}}_\exists s_2$ iff $\exists s_1 \in \rhd_\eta(s_2) \mid s_2 \succ^{t,\text{sure}}_\exists s_1 \wedge s_2 \not\succ^{t,\text{sure}}_\exists s_3 + s_1$.

Given $s_3 \in S_\eta, s_2 \in S_\eta, s_2 \notin \rightsquigarrow^{-1}_\eta(T^*), t \in T$: $s_3 \not\succ^{t,\text{sure}}_\exists s_2$ iff $\exists s_1 \in \rhd_\eta(s_2) \mid s_2 \succ^{t,\text{sure}}_\exists s_1 \wedge s_2 \succ^{t,\text{sure}}_\exists s_3 + s_1$.

F. Example about default arguments

s2 argues in favor of p against s1: s "le monde n'est pas fiable". s1 "le monde est fiable, bhl l'a dit". s2 "bhl est un clown, il s'est planté sur l'Irak". s3 "il avait raison sur l'Irak : l'Irak a des ADM". s4 "l'Irak n'a pas d'ADM, Bush l'a reconnu". Does s4 attack s3? "bhl est un clown, il s'est planté sur l'irak" + "l'irak n'a pas d'ADM, Bush l'a reconnu" VS "il avait raison sur l'Irak : l'Irak a des ADM" !

Measure problem?

G. Alternative definitions of finite defense

Define Finite defense-$\succ_\exists$-$\succ_\exists$-$\rhd_\eta$-dec as: $\succ_\exists^{-1}(s) \subseteq \succ_\exists(S_\eta \setminus \mathrm{im}(\rhd_\eta)) \Rightarrow \succ_\exists^{-1}(s) \subseteq \succ_\exists(S)$. Finite defense-$\succ_\exists$-$\succ_\exists$-$\rhd_\eta$-dec is insufficient to provide $T_\eta = T_i$. Define $s' \rhd_\eta^{\text{fail}} s$ iff $s' \rhd_\eta s \wedge \neg(s' \succ_\exists s)$. Consider $s_3 \succ_\exists s_2 \succ_\exists s_1$, $s_3' \succ_\exists s_2' \succ_\exists s_1$, and so on, and $s_4 \rhd_\eta^{\text{fail}} \{s_3, s_3', \ldots\}$. Then I really need infinitely many arguments to defend s_1 but Finite defense-$\succ_\exists$-$\succ_\exists$-$\rhd_\eta$-dec is artificially satisfied because the antecedent fails to trigger.

Define Finite defense-$\succ_\exists$-$\succ_\exists$-subsets as: $\succ_\exists^{-1}(s) \subseteq \succ_\exists(S) \Rightarrow \succ_\exists^{-1}(s) \subseteq \succ_\exists(S')$. Finite defense-$\succ_\exists$-$\succ_\exists$-subsets is insufficient to provide $T_\eta = T_i$. This is because Reinstatement allows for new attacks in $\succ_\exists$ (it only forbids new attacks in $\rhd_\eta$), thus we can forever transform previously failing attacks to new attacks, hence always satisfying Finite defense (always finite cover of $\succ_\exists$, but infinite cover of $\rhd_\eta$) but still not converging. Consider $s_3 \succ_\exists s_2 \succ_\exists s_1$, $s_3' \succ_\exists s_2' \rhd_\eta^{\text{fail}} s_1$, and so on; and $s_3' \succ_\exists s_2' \succ_\exists s_3 + s_1$, $s_3'' \succ_\exists s_2'' \rhd_\eta^{\text{fail}} s_3 + s_1$, and so on.

Define Finite defense-$\rhd_\eta$-$\succ_\exists$-startdec as: $\rhd_\eta^{-1}(s) \subseteq \succ_\exists(S_{\text{decisive}}) \Rightarrow \rhd_\eta^{-1}(s) \subseteq \succ_\exists(S)$. Finite defense-$\rhd_\eta$-$\succ_\exists$-startdec is insufficient to provide $T_\eta = T_i$. Consider $s_3 \succ_\exists s_2 \succ_\exists s_1$, $s_3' \succ_\exists s_2' \succ_\exists s_1$, and so on, and $s_5 \rhd_\eta^{\text{fail}} s_4 \rhd_\eta^{\text{fail}} s_1$. Then I really need infinitely many arguments to defend s_1 but Finite defense-$\rhd_\eta$-$\succ_\exists$-startdec is satisfied as there is no cover of the $\rhd_\eta$ attacks to s_1.

Define Finite defense-$\rhd_\eta$-$\rhd_\eta$-startdec as: $\rhd_\eta^{-1}(s) \subseteq \rhd_\eta(S_{\text{decisive}}) \Rightarrow \rhd_\eta^{-1}(s) \subseteq \rhd_\eta(S)$. Finite defense-$\rhd_\eta$-$\rhd_\eta$-startdec is insufficient to provide $T_\eta = T_i$. Consider $s_3 \succ_\exists s_2 \succ_\exists s_1$, $s_3' \succ_\exists s_2' \succ_\exists s_1$, and so on, and $s \rhd_\eta^{\text{fail}} \{s_2, s_2', \ldots\}$. Then I really need infinitely many arguments to defend s_1 but Finite defense-$\rhd_\eta$-$\rhd_\eta$-startdec is artificially satisfied because of s. Define Finite defense-$\rhd_\eta$-$\rhd_\eta$-subsets as: $\rhd_\eta^{-1}(s) \subseteq \rhd_\eta(S) \Rightarrow \rhd_\eta^{-1}(s) \subseteq \rhd_\eta(S')$ (for any $S \subseteq S_{\text{decisive}}$). Finite defense-$\rhd_\eta$-$\rhd_\eta$-subsets is (rightly) non satisfied in this example.

Define Finite defense-$\succ_\exists$-$\succ_\exists$-subsets as: $\succ_\exists^{-1}(s) \subseteq \succ_\exists(S) \Rightarrow \succ_\exists^{-1}(s) \subseteq \succ_\exists(S')$. Finite defense-$\rhd_\eta$-$\succ_\exists$-subsets $\not\Rightarrow$ Finite defense-$\succ_\exists$-$\succ_\exists$-subsets. Consider $s_2 \rhd_\eta^{\text{fail}} s_1$ (to be continued…)