How a new understanding alone can change preferences

Franz Dietrich
(LSE & Maastricht Univ.)

and

Christian List
(LSE)

Work in progress

LSE-Groningen Workshop
February 2009
London

Introduction

The phenomenon to be explained/modelled: s.o.'s preference change ('PC')

Standard explanation/model of PC (in formal decis. th.): new *information*

- Many PCs are indeed information-driven
  - e.g. wanting to buy the car after learning its price
- Many others aren't just information-driven, or so we think.
  - in the foll. examples, you're invited to ask yourself: was this just info?

Examples

- An Iraki initially wants (and intends) to stay peaceful, but during the war develops a preference for doing 'terrorism'.
- Preference to stay a single reverses after falling in love.
- Preference to try a drug *only once* changes after a first try.
- Preference to become a French cook is lost after visiting a slaughterhouse.
- Preference to get rich is lost after moving to a poor country.
- Preference switches from juice to wine after reading the menu.
- Preference for an early night disappears after drinking wine.
- Preference for getting up early disappears once the alarm clock rings.

Abstract and qualitative understanding

In these examples, a driving force is the gain of a qualitative understanding of certain propositionss

- like 'there is war', 'I get married', 'Animals are slaughtered', 'I got up at 7 am', ...[1]

S.o. may have an abstract w/o a qualitative understanding:

- Kids presumably have an abstract understanding of romantic love ('love is when there is kissing', a kid might say without a qualitatively understanding).

[1] Qualitative understanding of propositions might be traced back to qualitative understanding of *objects*, like war, love, animals' pain, tiredness, ...

Abstract and qualitative understanding

By definition, we say that someone

- **abstractly** understands a proposition if he conceptualises it, i.e. has some mental representation of it
- **qualitatively** understands a proposition A if he has imaginative access to its qualitative character; in short: if he can *imagine its quality*.[2]

  –> more on quality later

[2]Presumably, a qualitative understanding presupposes an abstract one.

Abstract and qualitative understanding

More examples

- Most people abstractly understand 'it is war' (conceptualised, say, in terms of people shooting at each other), but a qualitative understanding presumably requires having experiencing a war, or something similar.
- In Frank Jackson's thought experiment, Mary in her black-and-white chamber has an abstract understanding of 'applies are red' (conceptualised in terms of physical properties), but develops a qualitative understanding only after being shown an apple.[3]

Convention: When we say 'understanding' simpliciter, we refer to qualitative understanding (not just abstract understanding).

[3] Jackson goes on to reject physicalism. This quesion is not at stake here.

Two disclaimers

1. Our understanding-driven PC model
   - isn't meant to deny that *some other* PCs are info-driven;
   - could be combined with the standard rational-choice model, giving a model in which both understanding- and info-driven PCs can occur

–> So we don't aim to replace the standard model, but to complement it, so as to adequately cope with diff. kinds of PCs

2. Understanding-driven PCs
   - admittedly involve information-learning, namely learning a quality[4] (is this why they are often seen as a form of standard info-driven PCs?)
   - but go beyond pure information-learning[5]

[4]Provided we are realists about qualities; see the (metaphysical) Problem 4 raised later.

[5]See Problem 1 raised later.

Understanding- vs. information-driven PC

The core difference betw. these kinds of PC is one betw.

- learning to imagine the quality, say Q, of a proposition;
- removing an uncertainty, e.g. about whether a proposition truly has an (*already imagined*!) quality Q.[6]

In Charles Dickens' Christmas Carol, stingy miserable Scrooge is shown by ghosts different scenes of life; he thereby realises the quality of various propositions about life ('all celebrate together', ...) and becomes a generous man who wants to share his fortune. Scrooge's PC was

- understanding-driven if he had previously lacked *imagination* of the (beautiful) quality Q of 'all celebrate together'
- information-driven if he could previously imagine the quality Q but just didn't *believe* that 'all celebrate together' has this quality.

[6]The indiv. might imagine many qualities Q, Q', Q", ..., not knowing which one holds.

Understanding- vs. information-driven PC (ctd.)

Information-driven PC (in the standard model)

The indiv. can be in different *states of information* I

I is a set of currently possible worlds

I induces a preference order $\succeq_I$ (over the alternative in question)

Change in information state I implies change in preference $\succeq_I$

This reduces preference dynamics to belief dynamics

Understanding-driven PC (in our model)

The indiv. can be in different *states of understanding* S

S is a set of currently understood propositions

S induces a preference order $\succeq_S$ (over the alternatives in question)

Change in understanding S implies change in preference $\succeq_S$

This reduces preference dynamics to understanding dynamics

Understanding- vs. information-driven PC: a behavioural difference

Information-driven PC (in the standard model)

- Precludes dynamic inconsistency (good or bad?):
  - *stable* preference over sufficiently refined outcomes
  - PC only over non-refined outcomes (non-atomic events)
- Hence precludes commitment behaviour
  - i.e. precludes 'discarding an option to avoid taking it later'
  - [unless one allows Gul-Pesendorfer-type preferences that imply commitment behaviour w/o dynamic inconsistency]

Understanding-driven PC (in our model)

- May imply (explain) dynamic inconsistency:
  - PC over *however refined* outcomes
- Hence may imply (explain) commitment behaviour

Model of understaonding-driven PC

X: set of (mutually exclusive) **possibilities**, the objects of preference, e.g.:

- the indiv.'s choice alternatives (or decision paths)
- states of nature (not influenced by the indiv.)
- outcomes influence by many individuals (and perhaps nature)

Subsets of X: **propositions** or **events**

Model of understaonding-driven PC (ctd.)

$\mathbf{A}$: a (finite) set containing *certain* propositions $A \subseteq X$

- interpreted as the basic[7] propositions of interest,
  e.g. $\mathbf{A} = \{$'there is war', 'I am in love', 'it is freezing', 'I am clinically depressed', 'No one likes me'$\}$
- the propositions in $\mathbf{A}$ are called the **properties**
  [a slightly unusual use of the term 'property']
- A possibility $x \in X$ *has* a property $A \in \mathbf{A}$ if $x \in A$.

[7]No need to include compositions like 'there is peace *and* it is freezing'

Model of understaonding-driven PC (ctd.)

A **state of understanding** (or simply an **understanding**) is a set $S \subseteq \mathbf{A}$ of (currently understood) properties

Examples:

- a soldier in a war might have current understanding $S = \{$'there is war', 'it is freezing'$\}$
- a teenager might have current understanding $S = \{$'No one likes me', 'I am in love'$\}$
- a blind person's understanding excludes 'the sky is deep blue'

For each state of understanding $S \subseteq \mathbf{A}$, let $\succeq_S$ be the indiv.'s preference order on X in state of understanding S.[8]

[8]Throughout, a *(preference) order* (on a set) is a transitive complete binary relation (on the set). Derived relations: $\succ_S$ (strict preference) and $\sim_S$ (indifference).

Model of understaonding-driven PC (ctd.)

An objection: How can s.o. hold preferences between possibilities without understanding all of their properties?

Reply: By having an abstract understanding/conceptualisation.

- a child may differentiate (and hold a preference) betw. whether or not his parents have (romantic) love for each other, w/o a qualitatively understanding.
- Mary (in Jackson's thought experiment) has a concept of (and possibly preferences about) the world w/o qualitatively understanding its colour properties.

(I've a slide with a more sophisticated reply for the case that *even abstract understanding* can fail.)

Model of understaonding-driven PC (ctd.)

Two examples

Let each property $A \in \mathbf{A}$ have a 'value' $v(A) \in \mathbf{R}$.

- 'I'm in love' might have a high value, 'there is war' a low (negative) value, ...

Let preference $\succeq_S$ maximise one of the following 'utility' functions:

- $u(x) = \sum_{A \in \mathbf{A}: x \in A} v(A)$ (sum-total value of properties)
  here, $\succeq_S$ is the same at all understandings S
  so, gaining or losing understanding doesn't change preference
- $u_S(x) = \sum_{A \in S: x \in A} v(A)$ (sum-total value of *understood* properties)
  here, gaining understanding of property $A \in \mathbf{A}$
  - doesn't change utility of possibilities x w/o property A
  - changes utility of possibilities x with property A by adding $v(A)$

Understanding A with or w/o an information that A

The preference orders $(\succeq_S)_{S\subseteq \mathbf{A}}$ represent preference dynamics driven by gains and losses of understanding of properties in $\mathbf{A}$.
Gaining the understanding of a property A...
may happen in combination with (be triggered by) learning that A:

- e.g. gaining understanding of 'there is a thunderstorm' by hearing a thunder outside
- N.B.: not just the state of understanding, but also the state of *information* changes (from whatever it was before, say $I \subseteq X$, to $I \cap A$).

but could happen without learning that A:

- e.g. gaining understanding of 'there is a thunderstorm' by watching a movie involving a thunderstorm
- e.g. gaining understanding of 'my neighbour is in love' by falling in love oneself.
- N.B.: no change in information state.

We need conditions on preferences...

because
(i) not all patterns of preferences across states are plausible
(ii) we aim at a parsimonious representation of preferences
–> understanding-driven PC should be able to 'compete' in simplicity with information-driven PC

We will reach the parsimony goal (ii)

To compare:

- Bayesian belief-revision is 'simple' in that the beliefs $\mathsf{Pr}_I(.)$ across info states I are induced by a single (prior) belief $\mathsf{Pr}(.)$: for each I, $\mathsf{Pr}_I(.) = \mathsf{Pr}(.|I)$
- Standard info-driven PC is 'simple' in that the preferences $\succeq_I$ across info states I are induced by a single belief $\mathsf{Pr}(.)$ and utility fn. U: for all I, $\succeq_I$ is given by the expectation of U w.r.t. $\mathsf{Pr}(.|I)$.

Similarly,

- Our understanding-driven PC model will, via two conditions, become 'simple' in that the preferences $\succeq_S$ across states of understanding S are induced by a single 'ground order'.

First postulate

Postulate 1: 'Only the understood motivates'. The indiv. is indifferent between possibilities with the same understood properties. Formally: for every state of understanding $S \subseteq \mathbf{A}$ and possibilities $x, y \in X$, if $\{A \in S : x \in A\} = \{A \in S : y \in A\}$ then $x \sim_S y$.[9]

In other words, the indiv. is indifferent betw. possibilities of which he has the same understanding, where his *understanding of a possibility* x (in state of understanding S) is

$S_x = \{A \in S : x \in A\}$ (the set of understood properties of x).

Objection: What if the individual cares about propositions outside $\mathbf{A}$?

Reply: Then $\mathbf{A}$ was mis-specified. Simply augment $\mathbf{A}$!

[9] Given our assumption that each $\succeq_S$ is a weak order, this is equivalent to *independence of non-understood properties*: the preference between two states depends only the understood properties, i.e., for all $S \subseteq \mathbf{A}$ and all $x, x', y, y' \in X$, if $\{A \in S : x \in A\} = \{A \in S : x' \in A\}$ and $\{A \in S : y \in A\} = \{A \in S : y \in A\}$ then $x \succeq_S y \Leftrightarrow x' \succeq_S y'$.

First postulate (ctd.)

Objection: Even w/o understanding (romantic) love, a child may want to later be in love, as he realises how happy love makes.

Reply: What motivates the child here isn't romantic love but resulting happiness, *which he already understands*.

$->$ so Postulate 1 doesn't apply, as the two options x, y betw. which the child is non-indiff. differ in an understood property.

Counter-reply: Also the adult is motivated not by love but by resulting happiness. Preference isn't affected by which properties in $\mathbf{A}$ are understood, since properties matter only as means for a single (always understood!) goal: happiness.

Counter-counter-reply: We doubt existence of a single goal. It is unclear what this goal ('happiness'?) would consist in. A plural theory of motivation seems more plausible. (Literature in support. E.g. my welfarism paper, Griffin? and???)

First postulate (ctd.)

Objection: While the child doesn't yet understand love, he predicts that as an adult he will want to be in love. This makes him already now want his future self to be in love.

Reply: What motivates the child here is again not his future in-lovedness but own-future-preference-satisfaction, which he already now understands.

–> So again Postulate 1 doesn't apply (as we don't have $x_S = y_S$).

Counter-objection: One's present concern about the future is always about satisfying the future preferences.

Counter-counter-reply. We don't think so. Before s.o. gets drug addicted he disapproves of his future preferences (for regular drug consumption).

Two (other) interpretations of the state

We've so far interpreted S as containing the properties that are currently (qualitatively) understood.
–> Postulate 1 isn't obviously true, but a psychological hypothesis

There are (at least) 2 other interperetations of S (hence of our non-informational PC model):
(1) Informational simplification (a decision heuristic)
(2) Limited conceptualisation (a representational limitation)
–> Postulate 1 becomes true *by definition* under these alternative interpretations

(1) Informational simplification

- When forming his preferences $\succeq_S$ the agent neglects the properties $A \notin S$
- One (of many possible) decision heuristics
- Rationalisable by appeal to saving decision costs
- Postulate 1 holds by construction of $\succeq_S$.

(2) Limited conceptualisation

- The possibilities x in X are the *modeller's* possibilities.
- The agent's possibilities (in his conceptualisation) are coarser –> they are (in state $S \subseteq D$) representable as elements of $\{0,1\}^S$ rather than of X
- Why then define $\succeq_S$ as an order on X rather than $\{0,1\}^S$? –> $\succeq_S$ is the modeller's shorthand for the agent's preference order on $\{0,1\}^S$: $x \succeq_S y$ stands for $x_S \succeq^*_S y_S$ (where $\succeq^*_S$ is the agent's preference order over $\{0,1\}^S$, and $x_S \in \{0,1\}^S$ is given by $x_S(A) = 1 \Leftrightarrow x \in A$, and similarly for y_S).
- Again, Postulate 1 is true by construction of the orders $\succeq_S$.

Second postulate

While Postulate 1 addresses the structure of preference in a given state, the new postulate addresses change in preference after change in state.

Postulate 2: For every state of understanding $S \subseteq \mathbf{A}$, every possibilities $x, y \in X$, and every non-understood property $A \in \mathbf{A} \backslash S$ that holds in neither x nor y, $x \succeq_S y \Leftrightarrow x \succeq_{S \cup \{A\}} y$.

Example: s.o.'s preference between eating fish without wine and eating meat without wine doesn't change if he gains understanding of wine.

Second postulate (ctd.)

Postulate 2 (repeated): For every state of understanding $S \subseteq \mathbf{A}$, every possibilities $x, y \in X$, and every non-understood property $A \in \mathbf{A} \backslash S$ that holds in neither x nor y, $x \succeq_S y \Leftrightarrow x \succeq_{S \cup \{A\}} y$.

Objection: If I gain understanding of property A, the already understood properties of possibilities x and y may appear in a new light, changing preference between x and y.
E.g., suppose possibilities are dinner plans. After gaining understanding of wine, preference might change even between dinner plans x, y that don't include wine, because x and y might appear in a new light.

Second postulate (ctd.)

Reply: To handle this dinner example, there are 2 possibilities:

- If the indiv. has acquired a 'new' understanding of already understood properties (of main courses), then
  - our model as it stands can't represent this because understanding is represented as an on-off attitude;
  - but an extended model that allows for more than one 'way of understanding' can represent this, and the dinner example doesn't clash with the (suitably extended) Postulate 2.
- Under another interpretation of the dinner example, the PC was triggered not by a 'new' understanding of already understood properties but by the gain of understanding not just of A but also of other properties that are held by x or y. Then there is no conflict with Postulate 2 (which assumes that just one property A is newly understood).

A richness assumption

Richness In Possibilities. For every selection of (affirmed or negated) properties $A^* \in \{A, A^c\}$, $A \in \mathbf{A}$, there is a possibility $x \in X$ in which all A^*, $A \in \mathbf{A}$, hold.

Interpretation

- Recall that the properties in $\mathbf{A}$ are '(semantically) basic', presumably representing atomic propositions like 'the indiv. is in love', 'there is war'
- It is plausible that such basic properties are (intuitively) logically independent:
  - any combination of (affirmed or negated) properties is meaningful
- X is required to accommodate any such logical possibility

A richness assumption (ctd.)

Interpretation (ctd.)

- A *logical* possibility may be more theoretic than real:
  - 'there is war' and 'I am happy' might never go together
- Technically, a *logical* possibility may be

  - known not to hold, i.e. epistemically impossible (assuming X contains states of nature)

  - infeasible (assuming X contains choice alternatives)

N.B.: It is meaningful and perfectly imaginable to hold preferences $\succeq_S$ on a domain containing purely logical possibilities

$->$ Savage's and vNM's decision theories have their own richness assumptions on the preference domain

$\succeq_S$ can always be restricted back to the set of actual possibilities

An implicit richness assumption

Another richness assumption is **implicitly** built into the model:

- Every set of properties $S \subseteq \mathbf{A}$ is possible as a state of understanding.

Objection: One might argue that

- certain properties in $\mathbf{A}$ (i.e. certain basic propositions the indiv. may care about) are *always* or *never* understood
  - e.g. 'it is freezing' is *always* (qualitatively) understood
- understanding certain properties entails (not) understanding certain others
  - e.g. understanding 'it is war' precludes understanding 'there is harmony'
  - e.g. understanding 'there is harmony' entails understanding 'there is peace'

Reply: If a state of understanding $S \subseteq \mathbf{A}$ is indeed 'infeasible', $\succeq_S$ might be interpreted via a thought experiment.

... admittedly, that might be problematic.

Theorem

Theorem. Assume Richness in Possibilities. Then Postulates 1 and 2 hold if and only if there exist an order on the set $\mathcal{P}(\mathbf{A})$ of property sets, denoted $\succeq$ and called the *ground order*, that induces each $\succeq_S$ in the sense that $x \succeq_S y \Leftrightarrow \{A \in S : x \in A\} \succeq \{A \in S : y \in A\}$ for all possibilities $x, y \in X$.

Informally: whatever the current state of understanding S, the current preference $\succeq_S$ is induced by a single fixed ground order $\succeq$ in the sense that a world x is (currently) preferred to another y if and only if $\succeq$ ranks x's (currently) understood property combination above y's.

Remodelling understanding informationally?

The obvious attempt to construe understanding A informationally is to identify it with knowing that 'A has quality Q', where 'Q' is a name that refers to A's quality.
(–> The event 'A has quality Q' can't be represented in X, but in a a refined space.)

Remodelling understanding informationally?

Problem 1: neglecting part of the story

Gaining understanding of A is a two-fold gain w.r.t. the event 'A has quality Q', involving

(i) learning that the event holds (the informational gain),

(ii) but also, more fundamentally, developing the imagination of the quality Q, hence the concept of 'A has quality Q' (the imagination/awareness/conceptualisation gain)

–> w/o imagining Q he can't conceptualise 'A has quality Q' (but can conceptualise A, by abstract understanding

An informational remodelling accounts only for (i) but neglects (ii)

–> but (ii) is motivationally relevant (or so Postulate 1 claims)

B.t.w.: The ex ante lack of conceptualisation implies a lack of **negative introspection**: the indiv. doesn't know that he doesn't know that A has quality Q.

–> can be spelled out formally using the knowledge correspondence

Remodelling understanding informationally?

Problem 2: ascribing beliefs on non-conceptualised propositions

The model ascribes the agent beliefs (subjective probabilities) of all events, including the event 'A has quality Q' which, as noted, the indiv. doesn't conceptualise (ex ante).

Remodelling understanding informationally?

Problem 3: misrepresenting lack of understanding?

It is seems natural to model (qualitative) understanding of A as subjective certainty of 'A has quality Q'.

But is it also natural to

- model *non*-understanding of A (i.e., by def., inability to imagine Q) as subjective uncertainty that 'A has quality Q'?
- I.e., model 'I can't imagine Q' by 'my belief that A has quality Q is just 0.374'? (Any belief other than 0.374 seems equally ad hoc.)
- I.e., take a representational limitation for an uncertainty?

–> a category mistake?

Remodelling understanding informationally?

Problem 4 (metaphysical): belief where there is no fact?

It isn't obvious that there exists a fact about whether or not 'A has quality Q'. (Some wouldn't even call 'A has quality Q' a proposition.)

Reason: The quality Q might be viewed as a feature just of the observer's subjective experience (qualia), not of the external world.

If 'A has quality Q' is neither true nor false, should we ascribe the indiv. a belief of it?

Potential replies:

Either insist that 'A has quality Q' has a truth value (realism).

Or 'subjectivise' this event by reading it, e.g., as 'A has *subjective* quality Q if the individual understands A'... which seems to have a truth value.

Remodelling understanding informationally?

Despite all these worries, let's now do the remodelling.

–> The remodel limits attention to a single property, $A \in \mathbf{A}$;

–> See complementary slides for the general remodel.[10]

Refine each possibility $x \in X$ by splitting it into two (sub)possibilities:

- $(x, 1)$: represents 'x holds *and* A has quality Q'
- $(x, 0)$: represents 'x holds *and* A does not have quality Q'

where 'Q' denotes the newly understood quality of A (e.g. of 'I have a depression').

- New set of possibilities: $\overline{X} := X \times \{0, 1\} = \{(x, 0), (x, 1) : x \in X\}$.
- *Understanding* A is identified with *having the information* $\{(*, 1)\} = \{(x, 1) : x \in X\} \subseteq \overline{X}$ that A has quality Q
- The indiv.'s prior belief: a probability measure $\mathsf{Pr} : \mathcal{P}(\overline{X}) \rightarrow [0, 1]$
  –> which can be updated in light of information.[11]

[10] There, each possibility $x \in X$ is split not into 2 but into $2^{|\mathcal{A}|}$ (sub)possibilities.

[11] We assume that $\mathsf{Pr}(x_1) > 0$ for each $x \in X$. This is necessary to remodel the PC (specifically,

Remodelling understanding informationally?

The above worries can now be re-stated more formally:

Problem 1. The model can't distinguish between
- learning an event $E \subseteq \overline{X}$ (pure information),
- conceptualising *and* learning an event $E \subseteq \overline{X}$ (more than information).

So no distinction betw. information- and understanding-driven PC.

Problem 2. The space $\overline{X}$ reflects the modeller's representation of possibilities, not the indiv.'s (less refined) mental representation[12]. Yet the model ascribes beliefs within $\overline{X}$.

Problem 3. Is inability to imagine Q naturally modelled by Pr('A has quality Q') $= 0.374$?

Problem 4. What means the belief Pr('A has quality Q') if there isn't a fact about 'A has quality Q'?

to have all relevant conditional beliefs well-defined).

[12]Except when he is in the full-understanding state.

Remodelling understanding informationally?

Problem 5 (or perhaps just an observation)

A *loss* of understanding

- seems to us a (nearly) as important phenomenon as gain of understanding

  (–> As we grow older, do we gain or lose more understanding? Many youngsters complain about the lack of understanding of adults.)
- was easily modelled by removing A for the understanding state S.

Remodelling understanding informationally?

But loss of *information*

- is non-standard in decision theory;
- can't be modelled based on info. states.
  (–> if the indiv. is in an info state $I \subseteq \overline{X}$ in which he knows $E \subseteq \overline{X}$ (i.e. $I \subseteq E$), but then forgets E, what is his new info state? Presumably some state $I^* \subseteq \overline{X}$ s.t. $I \subseteq I^* \not\subseteq E$, but which exactly?)
- but could be modelled in a more complex (still 'informational') way (–> 'information sets' in game theory)

So: let's focus on remodelling *gain* of understanding of A.

Remodelling understanding informationally?

Let's go on with the informational model; the next steps are obvious:
1. As the indiv. learns event $\{(*, 1)\}$ that A has quality Q, his info changes from some initial state $I_0 \subseteq \overline{X}$ to new state $I := I_0 \cap \{(*, 1)\}$.
For simplic., info changes from $I_0 = \overline{X}$ ('no info') to $I = \{(*, 1)\}$ (A has quality Q).
2. Note: the belief (probability function) changes from $\mathsf{Pr}(.)$ to $\mathsf{Pr}(.|I)$.
3. Let $U : \overline{X} \to \mathbf{R}$ be a 'utility' function.
4. Events $C \subseteq \overline{X}$ are evaluated by their expected utility w.r.t. current belief.
So PC stems from change in exp. utility as belief changes from Pr to $\mathsf{Pr}(.|I)$.
5. We are particularly interested in PC betw. events of type $\{(x, *)\} = \{(x, 1), (x, 0)\} \subseteq \overline{X}$, which correspond to the possibilities $x \in X$ betw. which there is PC in our understanding-driven model.

Remodelling understanding informationally?

Problem 6: Postu. 1 is at odds with the informational model

To complete the 'remodel', we must translate Postu. 1-2.

Only part of Postulates 1-2 is relevant, because our (re)model focuses only on one fixed property $A \in \mathbf{A}$

–> A is the only property in $\mathbf{A}$ that may be non-understood

(–> see complementary slides for the general case, where any property in $\mathbf{A}$ can be non-understood)

The two postulates stated in the *original* (non-informational) model:

Postulate 1 (relevant part) For all $x, y \in X$ and all $S \in \{\mathbf{A}, \mathbf{A}\backslash\{A\}\}$, if $\{A' \in S : x \in A'\} = \{A' \in S : y \in A'\}$ then $x \sim_S y$.

Postulate 2 (relevant part) For all $x, y \in X$ such that $x \notin A$ and $y \notin A$, $x \succeq_{\mathbf{A}\backslash\{A\}} y \Leftrightarrow x \succeq_{\mathbf{A}} y$.

Remodelling understanding informationally?

To translate Postulates 1-2, we draw on the correspondence:

Original model	**Informational (re-)model**
X (set of possibilities)	$\overline{X}$ (set of refined possibilities)
$x \in X$	$\{(x, *)\} \subseteq \overline{X}$
understanding state S ($\mathbf{A}\backslash\{A\}$ or $\mathbf{A}$)	information $I_S \subseteq \overline{X}$ (defined below)
$x \succeq_S y$	$EU(\{(x, *)\} \cap I_S) \geq EU(\{(y, *)\} \cap I_S)$ ($\{(x, *)\}$'s expected utility conditional on information I_S is at least as high as $\{(y, *)\}$'s)
in particular...	
understanding state $\mathbf{A}\backslash\{A\}$ (A not understood)	information $I_{\mathbf{A}\backslash\{A\}} = \overline{X}$ (no information)
$x \succeq_{\mathbf{A}\backslash\{A\}} y$	$EU(\{(x, *)\}) \geq EU(\{(y, *)\})$
understanding state $\mathbf{A}$ (A understood)	information $I_{\mathbf{A}} = \{(*, 1)\} \subseteq \overline{X}$ (A known to have quality Q)
$x \succeq_{\mathbf{A}} y$	$U(x, 1) \geq U(y, 1)$

Remodelling understanding informationally?

Postulate 2 (relevant part; stated in the informational model) For all $x, y \in X$ such that $x \notin A$ and $y \notin A$, $EU(\{(x, *)\}) \geq EU(\{(y, *)\}) \Leftrightarrow U(x, 1) \geq U(y, 1)$.

A plausible condition! Because, whenever $x \notin A$ and $y \notin A$,

- by $x \notin A$ it is plausible that $U(x, 1) = U(x, 0)$, whence $EU(\{(x, *)\}) = U(x, 1)$,
- by $y \notin A$ it is plausible that, $U(y, 1) = U(y, 0)$, whence $EU(\{(y, *)\}) = U(y, 1)$.

Remodelling understanding informationally?

Postulate 1 (relevant part; stated in the informational model) For all $x, y \in X$ and each information I_S, $S \in \{\mathbf{A}, \mathbf{A}\backslash\{A\}\}$, if $\{A' \in S : x \in A'\} = \{A' \in S : y \in A'\}$ then $EU(\{(x, *)\} \cap I_S) = EU(\{(y, *)\} \cap I_S)$.

In other words: For all $x, y \in X$,

(i) if $\{A' \in \mathbf{A} : x \in A'\} = \{A' \in \mathbf{A} : y \in A'\}$ then $U(x, 1) = U(y, 1)$,

(ii) if $\{A' \in \mathbf{A}\backslash\{A\} : x \in A'\} = \{A' \in \mathbf{A}\backslash\{A\} : y \in A'\}$ then $EU(\{(x, *)\}) = EU(\{(y, *)\})$.

Postu. 1 (i.e. part (ii)[13]) is **implausible** from an orthodox angle $->$ for two reasons (see next slides)

[13]Part (i) is plausible, as we assume $\mathbf{A}$ to contain all properties that matter.

Remodelling understanding informationally?

Postu. 1, i.e. its part (ii), is **implausible** from an orthodox angle:

Reason 1. Imposing prior indifference between $\{(x, *)\}$ and $\{(y, *)\}$ is counterintuitive without the (unorthodox) appeal to lack of imagination, because:

- Mere uncertainty about A's quality doesn't imply indifference of whether property A holds...
  –> Expected-utility reasoning is sensitive to uncertain (dis)advantages.
  –> Uncertain qualities *can* motivate, non-imagined ones can't
- E.g., if x, y share the same properties $A' \in \mathbf{A} \backslash \{A\}$ (as in Postu. 1), x has property A but y doesn't, and the agent likes the quality Q (which property A *might* have), then presumably $EU(\{(x, *)\}) > EU(\{(y, *)\})$.

Remodelling understanding informationally?

Postu. 1, i.e. its part (ii), is **implausible** from an orthodox angle:

Reason 2. Imposing Postu. 1 *in the informational model* leads to pathological restrictions on the utility/probability functions U, Pr

... a kind of reductio ad absurdum

- of the informational model as a model of non-informational PC,
- i.e. of 'squeezing' non-informational PC (embodied by the essentially non-informational Postu. 1) into an informational revision setting.

Remodelling understanding informationally?

Reason 2 (ctd.)

One unnatural implication of Postu. 1: For all $x, y \in X$ that share the same properties in $\mathbf{A}$ except that x has A but y doesn't, if $U(x, 1) > U(y, 1)$ then $U(x, 0) > U(y, 0)$. (Proof: otherwise $EU(\{(x, *)\}) > EU(\{(y, *)\})$.)

In short: if property A (e.g. 'peace') is liked with quality Q (e.g. with a certain harmony) then it is disliked without Q.

'Why?', does one naturally ask.

Answer: 'That's what Postul. 1 implies in the informational model.'

Remodelling understanding informationally?

This leaves us with two alternatives.

1. **Not imposing** Post. 1 in the informat. model. Then the model
   - stays faithful to orthodox methodology and motivation
   - but doesn't remodel our understanding-driven model: preference can behave v. diff. (less pre-understanding indifference)
2. **Imposing** Postulate 1 (and 2). Then the informational model
   - does remodel the original model (to be precise, it isn't isomorphic to the original model but *encompasses* it[14])
   - but becomes non-orthodox and unnatural, because Postu. 1
     - goes against standard reasoning in decision th. under risk (*uncertain* qualities *do* motivate), while following a novel hypothesis (*non-imagined* qualities do *not* motivate)
     - leads to unnatural restrictions on utilities/probabilities

[14] I.e., it entails the same patterns of preferences $\succeq_S$ between elements $x \in X$ (identied with pairs $\{x_1, x_0\} \subseteq \overline{X}$) across understanding states S (identified with corresponding information states $I_S \subseteq \overline{X}$). But it entails more (desirable?) features, it is 'richer' (*strictly* encompassing). E.g., it states preferences *not only* betw. elements $x \in X$, i.e. pairs $\{x_1, x_0\} \subseteq \overline{X}$; and it entails dynamic consistency, while the original model is silent on this issue (cf. Problem 7).

Remodelling understanding informationally?

Problem 7 (or not a problem?)

By Pr('A has quality Q') < 1 (see a complementary slide), the informational model precludes ex-ante-certainty of future PC:

- e.g. pre-puberty certainty that puberty creates desire to meet women, or pre-menu-inspection certainty that menu-inspection creates craving for wine.
- reason: the info I (which triggers the PC) is ex ante uncertain to come.

Is this restriction desirable? That is:

- (under a normative interpretation of the model) Is an ex-ante certain future PC irrational?
- (under a descriptive interpretation) Is an ex-ante certain future PC impossible? We doubt this.

Complementary slides...

The case that even abstract understanding can fail

Our base-line interpretation is: the indiv. has an **abstract** understanding of all possibilities; only qualitative understanding varies.

Allow the indiv. to lack even abstract understanding of properties.

Our model can handle this.

But $\succeq_S$ needs re-interpretation.

Because the indiv.'s subjective representation of possibilities is coarser than the modeller's set of possibilities X.

Let X^* be a coarsening of X, containing 'subjective possibilities'.[15]

The modeller can represent the indiv.'s subjective preference order $\succeq_S^*$ on X^* uniquely as an order $\succeq_S$ on X:

$x \succeq_S y$ stands for $x^* \succeq^* y^*$, where x^* (y^*) is the subjective possibility that is a coarsening of x (y).

[15] Formally, one may view X^* as a partition of X into sets of (subjectively non-distinghished) possibilities. Subjective possibility $x^* \in X^*$ is a coarsening of modeller's possibility $x \in X$ iff $x^* \ni x$.

A generalisation: more than two ways of understanding

The model assumes that there is only one way of (qualitatively) understanding a property A:

More generally, one might

- consider a set W of 'ways of (qualitative) understanding', e.g.
  - $W = \{1\}$, as in our current model
  - $W = \{w^a, w^b\}$... two ways of understanding
  - $W = (0, 1]$, a continuum of 'levels', ranging from 0 (no qualitative understanding) to 1 (full qualitative understanding)
- re-define the indiv.'s state of understanding as a family $(w_A)_{A \in S}$:
  - S is, as before, a set $A \subseteq \mathbf{A}$ of understood properties;
  - w_A is the way in which $A \in S$ is understood.

Remodelling understanding informationally: what's the prior belief that A has quality Q?

The prior belief $\mathsf{Pr}(\{(*,1)\})$ that A has quality Q

(i) can't be 0 (otherwise the ex-post belief $P(.|\{(*,1)\})$ is undefined); so the indiv. ex ante *gives it a chance* that A has the (so far not even conceptualised!) quality Q;[16]

(ii) can't be 1 if there is to be a PC (otherwise $\mathsf{Pr}(.|\{(*,1)\}) = \mathsf{Pr}(.)$, i.e. no belief-change, so no PC); so the indiv. ex ante *isn't sure* that A has the quality Q.

The finding (i)-(ii) might disturb: on top of ascribing a belief of the *non-conceptualised* (Problem 2) and *perhaps non-factual* (Problem 4) proposition 'A has quality Q', this belief must be *non-trivial*.

- Any non-trivial belief (e.g. $\mathsf{Pr}(\{(*,1)\}) = 0.375$) seems ad hoc, perhaps more so than a non-trivial beliefs $P(\{(*,1)\}) \in \{0,1\}$.

[16] In fact, for each $x \in X$ we even need $\mathsf{Pr}(x,1) > 0$, to ensure well-definedness of x's ex-post expected utility $EU(\{(x,*)\} \cap \{(*,1)\})$ (which then reduces to the unconditional utility $U(x,1)$), hence of ex-post preferences.

Informational remodelling: general case

The following informational (re)model extends the earlier one in that it aims to represent *any* state of understanding $S \subseteq \mathbf{A}$ (rather than focussing on understanding-or-not of a *single* property $A \in \mathbf{A}$).

Refinded set of possibilities: $\widehat{X} := X \times \{0,1\}^{\mathbf{A}}$.

Refined possibility $(x, \mathbf{q}) \equiv (x, (q_A)_{A \in \mathbf{A}}) \in \widehat{X}$ represents

'x holds *and* $\begin{cases} \text{each property } A \in \mathbf{A} \text{ with } q_A = 1 \text{ has quality } \mathsf{Q}_A, \textit{ and} \\ \text{each property } A \in \mathbf{A} \text{ with } q_A = 0 \text{ doesn't have quality } \mathsf{Q}_A \end{cases}$',

where 'Q_A' denotes the quality of $A \in \mathbf{A}$ (the 'Brahms quality' if A is 'I'll visit a Brahms concert', a differ. quality if A is 'I have a depression', ...).

- *Understanding state* $S \subseteq \mathbf{A}$ is identified with the *information* $I_S := \{(*, \mathbf{q}) : \mathbf{q}_S = \mathbf{1}_S\} \subseteq \widehat{X}$ that each $A \in S$ has quality Q_A.

Informational remodelling: general case

Prior belief: a probability measure $\mathsf{Pr} : \mathcal{P}(\widehat{X}) \to [0,1]$.[17]
Posterior belief in state of understanding $S \subseteq \mathbf{A}$: $\mathsf{Pr}(.|I_S)$ $(= \mathsf{Pr}(.|\{(*,\mathbf{q}) : \mathbf{q}_S = \mathbf{1}_S\}))$.
$U : \widehat{X} \to \mathbf{R}$: a 'utility' function
Events $C \subseteq \widehat{X}$ are evaluated by their expected utility conditional on current information I_S, given by

$$\begin{aligned} EU(C \cap I_S) &= \sum_{(x,\mathbf{q}) \in \widehat{X}} U(x,\mathbf{q})\, \mathsf{Pr}(x,\mathbf{q}|C \cap I_S) \\ &= \frac{1}{\mathsf{Pr}(C \cap I_S)} \sum_{(x,\mathbf{q}) \in C \cap I_S} U(x,\mathbf{q})\, \mathsf{Pr}(x,\mathbf{q}). \end{aligned}$$

PC stems from change in exp. utility $EU(C \cap I_S)$ as information I_S changes, i.e. as S changes.
Of particular interest: the events $\{(x,*)\} \subseteq \widehat{X}$, $x \in X$, as these correspond to the possibilities $x \in X$ of the original model.[18]

[17] We assume that $\mathsf{Pr}(x,\mathbf{1}) > 0$ for each $x \in X$. This is necessary to remodel the PC (i.e. to have the relevant conditional beliefs well-defined).

[18] Event $\{(x,*)\}$'s conditional expected utility is given by $EU(\{(x,*)\} \cap I_S) = \frac{1}{\mathsf{Pr}(\{(x,\mathbf{q}):\mathbf{q}_S=\mathbf{1}_S\})} \sum_{\mathbf{q}:\mathbf{q}_S=\mathbf{1}_S} U(x,\mathbf{q})\, \mathsf{Pr}(x,\mathbf{q})$..

Informational remodelling: general case

Original model	**Informational (re-)model**
X (set of possibilities)	$\widehat{X} = X \times \{0,1\}^{\mathbf{A}}$ (set of refined poss.)
$x \in X$	$\{(x,*)\} \subseteq \widehat{X}$
understanding state $S \subseteq \mathbf{A}$ (all A in S understood)	the info $I_S = \{(*,\mathbf{q}) : \mathbf{q}_S = \mathbf{1}_S\} \subseteq \widehat{X}$ that each $A \in S$ has quality Q_A
$x \succeq_S y$	$EU(\{(x,*)\} \cap I_S) \geq EU(\{(y,*)\} \cap I_S)$ ($\{(x,*)\}$'s expected utility conditional on information I_S is at least as high as $\{(y,*)\}$'s)
	in particular...
state of understanding $\emptyset$	information $I_\emptyset = \widehat{X}$ (no information)
$x \succeq_\emptyset y$	$EU(\{(x,*)\}) \geq EU(\{(y,*)\})$
state of understanding $\mathbf{A}$ (*all* properties understood)	information $I_{\mathbf{A}} = \{(*,\mathbf{1})\} \subseteq \widehat{X}$ (*each* $A \in \mathbf{A}$ has quality Q_A)
$x \succeq_{\mathbf{A}} y$	$U(x,\mathbf{1}) \geq U(y,\mathbf{1})$

Informational remodelling: general case

Postulate 2 (stated in the informational model) For every $x, y \in X$, every information I_S, $S \subseteq \mathbf{A}$, and every property $A \in \mathbf{A} \backslash S$ that holds in neither x nor y, $EU(\{(x,*)\} \cap I_S) \geq EU(\{(y,*)\} \cap I_S) \Leftrightarrow EU(\{(x,*)\} \cap I_{S \cup \{A\}}) \geq EU(\{(y,*)\} \cap I_{S \cup \{A\}})$.

A plausible condition! Because, whenever $x \notin A$ and $y \notin A$,

- by $x \notin A$ it is plausible that $U(x, \mathbf{q})$ doesn't depend on q_A, whence $EU(\{(x,*)\} \cap I_S) = EU(\{(x,*)\} \cap I_{S \cup \{A\}})$ (by $I_{S \cup \{A\}} = I_S \cap \{(*, \mathbf{q}) : q_A = 1\}$),
- by $y \notin A$ it is plausible that, $U(y, \mathbf{q})$ doesn't depend on q_A, whence $EU(\{(y,*)\} \cap I_S) = EU(\{(y,*)\} \cap I_{S \cup \{A\}})$ (again by $I_{S \cup \{A\}} = I_S \cap \{(*, \mathbf{q}) : q_A = 1\}$).

Informational remodelling: general case

Postulate 1 (stated in the informational model) For every $x, y \in X$ and every information I_S, $S \subseteq \mathbf{A}$, if $\{A \in S : x \in A\} = \{A \in S : y \in A\}$ then $EU(\{(x, *)\} \cap I_S) = EU(\{(y, *)\} \cap I_S)$.

Postu. 1 is **implausible** from an orthodox angle:

Reason 1. Indifference betw. $\{(x, *)\} \cap I_S$ and $\{(y, *)\} \cap I_S$ is counterintuitive without appealing to a non-orthodox motivation (lack of imagination of the qualities of the properties not in S).[19]

Reason 2. Imposing Postu. 1 *in the informational model* leads to pathological restrictions on the utility/probability functions U, Pr.

[19] Except if $S = \mathbf{A}$. Then x and y have exactly the same properties in $\mathbf{A}$, whence the indifference is plausible since $\mathbf{A}$ is supposed to contain all relevant properties.

The standard model revisited

[I won't be 100% precise, partly in order to stay compatible with both
- standard modelling practice in decision or game theory
- different decision theories (Savage, Jeffrey)]

Alternatives of interest: $A, B, ...$
- e.g. the alternatives in our examples ('I behave peacefully', 'I do terrorist attacks', 'I get rich', 'I share with others')
- e.g. indiv.'s own actions, or events not under own control, ...

The standard model revisited (ctd.)

Fixed preferences $\succeq$ between *lotteries* L of (let me say) 'final outcomes'

- $\succeq$ is typically given by maximising the lottery's expected utility w.r.t. some fixed 'utility' function of final outcomes
- the final outcomes might be
  - the 'outcomes' in a decision problem under risk, or in a game
  - the 'outcomes' in von-Neumann-Morgenstern or 'consequences' in Savage or 'worlds' in Jeffrey.[20]

[20] In Savage and Jeffrey, $\succeq$ isn't a primitive but follows by maximising the exp. utility of final-outcome-lotteries w.r.t. a utility fn. whose existence a theorem provides. The primitive is rather a pref. rel. over Savage acts (in Savage) or events in a Boolean algebra (in Jeffrey).

The standard model revisited (ctd.)

Given an info state I (and prior beliefs), each alternative A leads to a lottery L_I^A over final outcomes

So we obtain, as a by-product:

Information-dependent preferences $\succeq_I$ between alternatives $A, B, ...$:
- $\succeq_I$ is given by $A \succeq_I B \Leftrightarrow L_I^A \succeq L_I^B$ for all alternatives A, B
-> an alternative is preferred to another iff it leads to a preferred lottery

The standard vs. new PC model: difference in ascribed mental state

The new model ascribes less sophistication to the individual:

- No need for highly refined 'final outcomes'
- No need to postulate a highly sophisticated indiv. who holds
  - preferences between *lotteries*
  - probabilistic belief attitudes, revised via Bayes' rule
- In fact, no need to invoke at all beliefs, information states, uncertainty, risk, lotteries
- But: we invoke states of *understanding*
- [The indiv. *might in addition* hold probabilistic beliefs, an information state, preferences over lotteries
  –> then he can in addition have informational PC's]