Defining Deliberated Choice and Theories Thereof

Olivier Cailloux

LAMSADE, Université Paris-Dauphine, PSL

29th May, 2024

Outline

1. Deliberated choice
2. Theories of deliberated choice
3. Properties and existence of theories
4. Discussion

Outline

Deliberated choice

- Individual i wonders about choosing some option among two possibilities
- Possible choices $\{\varphi, \neg\varphi, 0\}$ meaning "pick first option", "pick second option", "no preference"
- Examples: coke VS milkshake, vegan diet VS meat, increase inheritance tax, ...
- Shallow choice: the one without arguments
- Deliberated choice: the one that is stable facing counter-arguments
- Represents the choice after having considered all arguments from a given set of arguments

Formal context

- Options $P = \{\varphi, \neg\varphi, 0\}$
- Individuals I
- Arguments $\mathscr{A} = \{a_1, \dots\}$
- Behavior function $\rightsquigarrow$: the reactions of individuals to arguments (unknown but partially observable)

Example: inheritance tax

- Options $P = \{\varphi = \text{“increase”} = \mathit{incr}, \neg\varphi = \text{“do not increase”}, 0 = \text{“indifference”}\}$
- Individuals I: the persons in this room
- Arguments $\mathscr{A}$: a set of fifty arguments in favor or against increasing taxes (demographic facts, principles of justice. . .)
- Behavior function $\rightsquigarrow$: however the individuals react to the arguments

Sequence of arguments

- $\alpha \in \mathscr{A}^{<\mathbb{N}}$: a finite sequence of arguments
- $\rightsquigarrow_i \in P^{\mathscr{A}^{<\mathbb{N}}}$: $\alpha \rightsquigarrow_i \varphi$ iff individual i after seeing α (in order) opts for φ (also $\rightsquigarrow_i(\alpha) = \varphi$)
- Behavior function $\rightsquigarrow \in P^{\mathscr{A}^{<\mathbb{N}}I} = \{\rightsquigarrow_i \mid i \in I\}$

Example: behavior function

- $\emptyset \rightsquigarrow_{\text{Franz}} \textit{incr}$: Franz opts for *incr* without arguments
- $(a_1) \rightsquigarrow_{\text{Franz}} \neg\textit{incr}$: Franz rejects *incr* if given a_1
- $(a_1, a_2) \rightsquigarrow_{\text{Franz}} \textit{incr}$: Franz opts for *incr* if given a_1 then a_2
- $(a_1, a_2) \rightsquigarrow_{\text{Olivier}} \textit{incr}$, $(a_2, a_1) \rightsquigarrow_{\text{Olivier}} \neg\textit{incr}$: Olivier opts for *incr* if given a_1 then a_2 but not the other way around

$\rightsquigarrow$ encodes the reactions of all individuals to every possible sequence of arguments

Decisive argument

Decisive argument

a is *decisive* for i in favor of φ iff it convinces i whenever it appears within the last two arguments:

$$a \hookrightarrow_i \varphi \iff \forall \alpha \mid a \in \alpha_{[\![\#\alpha-1,\#\alpha]\!]} : \alpha \rightsquigarrow_i \varphi$$

Uniqueness

If a is decisive for i in favor of φ, there is no decisive argument for i in favor of any $p \neq \varphi$

Example: decisive argument

- Is a_1 decisive for Olivier?
- Is a_2 decisive for Franz?

Decisive argument

Decisive argument

a is *decisive* for i in favor of φ iff it convinces i whenever it appears within the last two arguments:

$$a \hookrightarrow_i \varphi \iff \forall \alpha \mid a \in \alpha_{[\![\#\alpha - 1, \#\alpha]\!]} : \alpha \rightsquigarrow_i \varphi$$

Uniqueness

If a is decisive for i in favor of φ, there is no decisive argument for i in favor of any $p \neq \varphi$

Example: decisive argument

- Is a_1 decisive for Olivier? No (not in favor of 0 or $\neg incr$ as $(a_1, a_2) \rightsquigarrow_i incr$ and not in favor of $incr$ as $(a_2, a_1) \rightsquigarrow_i \neg incr$)
- Is a_2 decisive for Franz?

Decisive argument

Decisive argument

a is *decisive* for i in favor of φ iff it convinces i whenever it appears within the last two arguments:

$$a \hookrightarrow_i \varphi \iff \forall \alpha \mid a \in \alpha_{[\![\#\alpha-1,\#\alpha]\!]} : \alpha \rightsquigarrow_i \varphi$$

Uniqueness

If a is decisive for i in favor of φ, there is no decisive argument for i in favor of any $p \neq \varphi$

Example: decisive argument

- Is a_1 decisive for Olivier? No (not in favor of 0 or $\neg incr$ as $(a_1, a_2) \rightsquigarrow_i incr$ and not in favor of $incr$ as $(a_2, a_1) \rightsquigarrow_i \neg incr$)
- Is a_2 decisive for Franz? Assuming that $(\dots, a_2) \rightsquigarrow_{\text{Franz}} incr$ and that $(\dots, a_2, .) \rightsquigarrow_{\text{Franz}} incr$, it is

Deliberated choice

Deliberated choice

The deliberated choice of i is p iff there is a decisive argument for i in favor of p; if no such $p \in P$ then it is $\emptyset$:

$$\begin{cases} \pi_i = p & \iff \exists a \mid a \hookrightarrow_i p \\ \pi_i = \emptyset & \iff \forall p \in P, \nexists a \mid a \hookrightarrow_i p \end{cases}$$

Example: deliberated choice

- π_{Franz}?

Deliberated choice

Deliberated choice

The deliberated choice of i is p iff there is a decisive argument for i in favor of p; if no such $p \in P$ then it is $\emptyset$:

$$\begin{cases} \pi_i = p & \iff \exists a \mid a \hookrightarrow_i p \\ \pi_i = \emptyset & \iff \forall p \in P, \nexists a \mid a \hookrightarrow_i p \end{cases}$$

Example: deliberated choice

- π_{Franz}? *incr*

Outline

At this stage

- Someone's deliberated choice π_i is well defined given $\leadsto$
- But we don't know $\leadsto$
- And we can't observe all of it!
- We need to phrase theories and determine how to validate them

Claims

Claim

A claim is a set $C \subseteq P^{\mathscr{A}^{<\mathbb{N}}I}$ of behavior functions $\rightsquigarrow$ considered as the possible ones

The claim excludes the complementary behaviors!

Example claims

- "Franz deliberately prefers *incr*" ($C = \{\rightsquigarrow \mid \exists a \mid a \hookrightarrow_{\text{Franz}} incr\}$)
- "Olivier never changes his mind given a_1"
  ($C = \{\rightsquigarrow \mid \forall \alpha : \rightsquigarrow_{\text{Olivier}}(\alpha) = \rightsquigarrow_{\text{Olivier}}(\alpha, a_1)\}$)
- "Olivier reacts exactly like Franz" [$\forall \alpha : \rightsquigarrow_{\text{Olivier}}(\alpha) = \rightsquigarrow_{\text{Franz}}(\alpha)$]
- Combinations of the above

Theories

Claim

A claim is *trivial* iff it contains all behaviors

$$C_{\text{trivial}} = P^{\mathscr{A}^{<\mathbb{N}} I}$$

Theory

A theory is a non trivial claim

Questions to be explored

- What should be postulated about observations? (Observable sets and Anonymity)
- What is a useful theory? (Indicativeness)
- How to ensure the correctness of a theory? (Falsifiability)

Observations

- We cannot "undo" exposure to arguments
- For a given i, we cannot observe both $\rightsquigarrow_i(a_1, a_2)$ and $\rightsquigarrow_i(a_3, a_4)$.
- We can only observe the reactions of i to sets of increasing sequences, such as $\langle(\emptyset), (a_3), (a_3, a_4), (a_3, a_4, a_1), \ldots\rangle$

Franzdoes not forget

- Assume that we observe that $(a_2) \rightsquigarrow_{\text{Franz}} incr$
- Now we cannot observe $(a_1) \rightsquigarrow_{\text{Franz}} \neg incr$
- We can only observe $(a_2, a_1) \rightsquigarrow_{\text{Franz}} incr$

- However, we can observe incompatible sequences on *different* individuals (e.g. $\rightsquigarrow_i(a_1, a_2)$ and $\rightsquigarrow_j(a_3, a_4)$)

Possible observations

- An observation is a set of triples $\theta \subset \mathscr{A}^{<\mathbb{N}} \times I \times P$
- The possible observations are the finite sets of triples $\theta \subset \mathscr{A}^{<\mathbb{N}} \times I \times P$ such that for a given i, the sequences of arguments related to i in θ forms an increasing sequence
- Let Θ denote that set of possible observations
- Let $\Theta \cap \mathscr{P}(\rightsquigarrow)$ denote the set of possible *observables*: observations that are compatible with $\rightsquigarrow$

Outline

Anonymity

Anonymity requires to not care about the identity of individuals

Anonymous theory

A theory T is anonymous iff it is closed under renaming of individuals:

$$\forall \sigma : I \leftrightarrow I, \rightsquigarrow \in T : (\rightsquigarrow \circ \sigma) \in T.$$

An anonymous theory does not distinguish individuals beyond their behaviors as captured by $\rightsquigarrow$ (informational constraint similar to Arrow's IIA).

Anonymity of theories

- "Olivier never changes his mind given a_1"?
- "Everybody opts for the same choice given a_1"?

Anonymity

Anonymity requires to not care about the identity of individuals

Anonymous theory

A theory T is anonymous iff it is closed under renaming of individuals:

$$\forall \sigma : I \leftrightarrow I, \rightsquigarrow \in T : (\rightsquigarrow \circ \sigma) \in T.$$

An anonymous theory does not distinguish individuals beyond their behaviors as captured by $\rightsquigarrow$ (informational constraint similar to Arrow's IIA).

Anonymity of theories

- "Olivier never changes his mind given a_1"? Not anonymous
- "Everybody opts for the same choice given a_1"?

Anonymity

Anonymity requires to not care about the identity of individuals

Anonymous theory

A theory T is anonymous iff it is closed under renaming of individuals:

$$\forall \sigma : I \leftrightarrow I, \rightsquigarrow \in T : (\rightsquigarrow \circ \sigma) \in T.$$

An anonymous theory does not distinguish individuals beyond their behaviors as captured by $\rightsquigarrow$ (informational constraint similar to Arrow's IIA).

Anonymity of theories

- "Olivier never changes his mind given a_1"? Not anonymous
- "Everybody opts for the same choice given a_1"? Anonymous

Informativeness and indicativeness

- A theory may fail to inform about anyone's deliberated choice (example?)
- A theory may inform only about numbers ("More individuals deliberately prefer *incr* than $\neg$*incr*")
- A theory may indicate something about someone's deliberated choice when knowing some of their reactions to arguments

Indicativeness

A theory T is indicative iff for some observations about i, i's deliberated choice, considering any behavior compatible with the observations and T, is a single $p \in P$

An indicative theory

"If i chooses *incr* given (a_1, a_2) then her deliberated choice is *incr*"

Informativeness and indicativeness

- A theory may fail to inform about anyone's deliberated choice (example? "Olivier never changes his mind given a_1")
- A theory may inform only about numbers ("More individuals deliberately prefer *incr* than $\neg$*incr*")
- A theory may indicate something about someone's deliberated choice when knowing some of their reactions to arguments

Indicativeness

A theory T is indicative iff for some observations about i, i's deliberated choice, considering any behavior compatible with the observations and T, is a single $p \in P$

An indicative theory

"If i chooses *incr* given (a_1, a_2) then her deliberated choice is *incr*"

Indicativeness

Example (An indicative theory)

"If i chooses *incr* given (a_1, a_2) then her deliberated choice is *incr*"

$$[\forall i \in I : (a_1, a_2) \rightsquigarrow_i incr \implies \pi_i = incr]$$

Validity

- So far: syntactic properties (can be checked without querying $\rightsquigarrow$)
- We need to check that the theory *holds*
- Holding is an empirical property

Holding

A theory T holds iff $\rightsquigarrow \in T$

Falsifiability

Falsifiability

A theory T is *falsifiable* iff whatever the real behavior function is, if it is not in T then we can observe that it is not:

$$\forall \rightsquigarrow \notin T : \Theta \cap \mathscr{P}(\rightsquigarrow) \nsubseteq \cup_{\rightsquigarrow' \in T} \mathscr{P}(\rightsquigarrow').$$

Falsifiability

- $[\forall i \in I : (a_1) \rightsquigarrow_i incr]$?
- Given i: $[(a_1) \rightsquigarrow_i incr \vee (a_2) \rightsquigarrow_i incr]$?
- $[\exists i \in I \mid (a_1) \rightsquigarrow_i incr]$?

Falsifiability

Falsifiability

A theory T is *falsifiable* iff whatever the real behavior function is, if it is not in T then we can observe that it is not:

$$\forall \rightsquigarrow \notin T : \Theta \cap \mathscr{P}(\rightsquigarrow) \not\subseteq \cup_{\rightsquigarrow' \in T} \mathscr{P}(\rightsquigarrow').$$

Falsifiability

- $[\forall i \in I : (a_1) \rightsquigarrow_i incr]$? Falsifiable
- Given i: $[(a_1) \rightsquigarrow_i incr \vee (a_2) \rightsquigarrow_i incr]$?
- $[\exists i \in I \mid (a_1) \rightsquigarrow_i incr]$?

Falsifiability

Falsifiability

A theory T is *falsifiable* iff whatever the real behavior function is, if it is not in T then we can observe that it is not:

$$\forall \rightsquigarrow \notin T : \Theta \cap \mathscr{P}(\rightsquigarrow) \not\subseteq \cup_{\rightsquigarrow' \in T} \mathscr{P}(\rightsquigarrow').$$

Falsifiability

- $[\forall i \in I : (a_1) \rightsquigarrow_i \mathit{incr}]$? Falsifiable
- Given i: $[(a_1) \rightsquigarrow_i \mathit{incr} \vee (a_2) \rightsquigarrow_i \mathit{incr}]$? Not falsifiable
- $[\exists i \in I \mid (a_1) \rightsquigarrow_i \mathit{incr}]$?

Falsifiability

Falsifiability

A theory T is *falsifiable* iff whatever the real behavior function is, if it is not in T then we can observe that it is not:

$$\forall \rightsquigarrow \notin T : \Theta \cap \mathscr{P}(\rightsquigarrow) \nsubseteq \cup_{\rightsquigarrow' \in T} \mathscr{P}(\rightsquigarrow').$$

Falsifiability

- $[\forall i \in I : (a_1) \rightsquigarrow_i \mathit{incr}]$? Falsifiable
- Given i: $[(a_1) \rightsquigarrow_i \mathit{incr} \vee (a_2) \rightsquigarrow_i \mathit{incr}]$? Not falsifiable
- $[\exists i \in I \mid (a_1) \rightsquigarrow_i \mathit{incr}]$? Not falsifiable iff I is infinite

A possibility theorem

Suitability

A theory T is *suitable* iff it holds and is anonymous, falsifiable and indicative.

With sufficent consensus, a suitable theory exists.

Situation admitting a suitable theory

If some argument is decisive for all individuals, then a suitable theory exists. (Formal condition: $\exists p \in P, a \in \mathscr{A} \mid \forall i \in I : a \hookrightarrow_i p$.)

An impossibility theorem

However, suitable theories generally do not exist.

Theorem (Situation admitting no suitable theory)

If some behavior admits a decisive argument for φ and some admits another decisive argument for $\neg\varphi$ and every behavior with some decisive argument is shared by infinitely many individuals and they all agree to start with, then no suitable theory exists.
(Formal condition: $\exists a_1 \neq a_2 \in \mathscr{A}, f_1, f_2 \in P^{\mathscr{A}^{<\mathbb{N}}} \mid a_1 \hookrightarrow_{f_1} \varphi \land a_2 \hookrightarrow_{f_2} \neg\varphi$ $\land \forall f, f' \in \rightsquigarrow(I) : f(\emptyset) = f'(\emptyset) \land \#\rightsquigarrow^{-1}(f) \notin \mathbb{N}$.)

Ongoing work: characterize those situations and search for workarounds!

Outline

Deliberated choice

- Deliberated choices complement shallow choices
- They retain some attractive features about shallow choices: observability, precision, choice semantics
- Formal definitions about deliberated choices permit to clarify concepts and compatibilities ("philosophers look for incompatibilities")
- Deliberated choices could constitute a legitimate basis for individual decision support
- Deliberated choices could constitute a legitimate basis for collective decision support

Normative VS empirical aspects

- Social choice theory separates normative choices (which properties one wants) from deductive aspects (which are compatible; what rule to use)
- This endeavor: separate the normative choice (the set of arguments, the protocol of observation, the desired properties of theories) from the empirical content (which theories are suitable, which arguments convince individuals)
- This approach may permit to frame some disagreements about action as empirical questions
- Long term goals: study sophisticated opinionated normative theories (Rawls, Nozick, Chomsky); apply to discuss nudging

Thank you for your attention!

Verifiability

Verifiability

- A theory T is verifiable in principle iff for some observations, T is deducible from the observations

$$\exists \theta \in \Theta \mid \forall \rightsquigarrow \in P^{\mathscr{A}^{<\mathbb{N}^{I}}} : (\theta \subset \rightsquigarrow \implies \rightsquigarrow \in T)$$

- A theory T is verifiable effectively iff for some observables, T is deducible from the observations

$$\exists \theta \in \Theta \cap \mathscr{P}(\rightsquigarrow) \mid \forall \rightsquigarrow \in P^{\mathscr{A}^{<\mathbb{N}^{I}}} : (\theta \subset \rightsquigarrow \implies \rightsquigarrow \in T)$$

Note that effective verifiability ensures that the theory holds. But:

Indicativeness and Verifiability are incompatible

When $\#\mathscr{A} \geq 2$, if T is indicative, then T is not verifiable

Falsifiability: an attempt

Falsifiability (attempt)

A theory T is *falsifiable* iff some observations permits to falsify it:

$$\Theta \not\subseteq \cup_{\rightsquigarrow' \in T} \mathscr{P}(\rightsquigarrow').$$

Fails!

An intuitively non falsifiable theory

- $(a) \rightsquigarrow_i \varphi \vee (a') \rightsquigarrow_i \varphi$ is not falsifiable (okay)
- $\alpha \rightsquigarrow_j \varphi \wedge [(a) \rightsquigarrow_i \varphi \vee (a') \rightsquigarrow_i \varphi]$ is falsifiable (should not be)