Lockean Beliefs, Dutch Books and Scoring Rules

Daniel Rothschild

University College London

LSE, 11 March 2018

Lockean Thesis

- On the Lockean thesis one ought to believe a proposition if and only if one assigns it a credence at or above a threshold
- Connects all-or-nothing beliefs to credences.

Foley 1992

Conceptual Motivations

- Preface paradox i.e. the rationality of inconsistency.
- Requiring probability 1 seems too strong, but we need some connection, as graded belief and all or nothing beliefs are not different systems.
- Primacy of graded beliefs. (?)

Linguistic Arguments for Lockeanism

- Neg-raising and weakness.

(1) Bill doesn't think it will rain.
$\rightsquigarrow$ Bill thinks it won't rain.

- Linguistic judgments.

(2) Zelda isn't sure it's going to rain, but she (still) thinks it will.

(3) Who does Olga think/believe won the election?
$\not\rightsquigarrow$ There is someone Ogla is sure won the election.

Hawthorne, Rothschild, and Spectre 2016, Rothschild 2019

Problems

- Allows inconsistent beliefs.
- Allows failure of closure.
- Hard to find the right threshold.

Leitgeb 2014 gives a non-trivial version of Lockeanism that ensures closure.

Foundations of non-Lockean Belief

- Full beliefs (or beliefs with probability 1) are closed, consistent sets as characterized by propositional logic.
- Graded belief satisfying axioms of probability can be captured by Dutch Book arguments or accuracy arguments.

Hintikka 1962; Stalnaker 1970, Finetti 1974; Joyce 1998; Pettigrew 2016; Ramsey 1926

Foundations of Lockean Belief

- The definition of Lockeanism:
  believe p iff only if you assign p a credence greater than t.
- But can we give an independent characterization of what it is to be a Lockean agent?

Consistency Characterizations

- If threshold is $> \frac{1}{2}$ then we cannot believe P and $\neg P$, on the Lockean thesis.
- But merely being pairwise consistent is not sufficient for being Lockean.
- In other words, the following two statements are not equivalent:[1]
  - A set of propositions, B, is pairwise consistent.
  - There is a probability function that assigns each member of B a probability greater than $\frac{1}{2}$

[1] Trivially since Lockean sets are single-premise closed. But even adding this condition is not enough. Example: $W = \{1, 2, 3, 4, 5\}$, this collection of beliefs set is pairwise consistent but not Lockean for $t > .5$: $\{p : p \supseteq \{1, 2, 3\}$ or $p \supseteq \{1, 4, 5\}\}$.

Independent Foundations

- Here, the goal is to provide independent characterizations of Lockeanism.
- I do so with:
  - Dutch Book Argument
    (novel arguments, though some of the mathematical results have analogues in literature on imprecise probabilities)
  - Accuracy argument
    (advances current literature by providing more general results and proving Easwaran's conjecture.)

Connections to Foundational Projects

- Graded beliefs first (e.g. Dorst).
- Full beliefs first (e.g. Easwaran).
- Bets first (e.g. classic decision theory).

Dorst 2019; Easwaran 2016; Finetti 1974; Ramsey 1926

Lockeanism Made Precise

- Set of worlds W, enumerated, $w_1 \ldots w_m$
- Set of propositions, $P = 2^W$, enumerated $p_1 \ldots p_n$
- Set of propositions believed $B \subseteq P$ enumerated $b_1 \ldots b_o$
- t a real-valued threshold $0 \leq t \leq 1$

Probabilities

- Credences over P:
  function $c : P \to [0,1]$ s.t. $c(W) = 1$, $c(\emptyset) = 0$ and for any $X \subseteq W$, $c(X) = \sum_{w \in W} c(\{w\})$.
- In vector terms, treat $\mathbf{c}$ as a m-dimensional non-negative vector whose coordinates sum to 1.
- Correspond to a proposition p, the m -dimensional vector $\mathbf{p}$ s.t. $p_i = 1$ if $\mathbf{w_i} \in p$ and $p_i = 0$ otherwise.
- Then $c(p) = \mathbf{c} \cdot \mathbf{p}$

Which Sets of Beliefs are Lockean

- **B** is Lockean compatible relative to t iff there is a credence function c, such that $c(b) \geq t$ (or $\mathbf{c} \cdot \mathbf{b} \geq t$) for all $b \in B$.
- **B** is Lockean complete iff $c(p) \geq t$ iff $p \in B$ for all $p \in P$
- **B** is almost Lockean complete iff if $c(p) > t$ then $p \in B$ and if $c(p) < t$ then $p \notin B$.

Dutch Book Arguments

- What is the Dutch Book argument?
- What kind of betting behavior should Lockeanism license?

Traditional Dutch Book Argument 1/2

- Bets on propositions: return \$1 if propositions is true and \$0 if false.
- Obviously everyone should be willing to accept a free bet for or sell such a bet for \$1.
- But we might assume everyone has some indifference point \$$x$ at which they are willing to either buy or sell bets.

Finetti 1974; Ramsey 1926

Traditional Dutch Book Argument 2/2

- We can call the indifference points for all the propositions, their numerical confidences.
- We can call a book of bets a collection of instructions to either buy or sell bets on propositions at the numerical confidence.
- A Dutch Book is book of bets that guarantees a loss at every world.
- Standard Dutch Book theorems show (roughly) that a set of numerical confidences over P are probabilistically coherent iff they are not subject to a Dutch Book.

How Should Lockeans Bet?

- Since Ramsey, graded beliefs have been associated with betting dispositions.
- What about all-or-nothing beliefs?
- First hypothesis: a Lockean should be willing to pay $\$t$ (or less) for a bet on any proposition in B.[2]

[2] We can extend system to allow t to vary with each proposition.

One-Way Dutch Book 1/2

- Take an agent with beliefs B and threshold t.
- We can use a non-negative o-dimensional vector $\mathbf{x}$, the stake vector, to represent the number of bets the agent buys in each proposition $b_1 \dots b_o$ (i.e. the agent buys x_i bets in b_i).
- At world w this leads to payoff:

$$\sum_{1 \leq i \leq o, w \in b_i} x_i(1-t) + \sum_{1 \leq i \leq o, w \notin b_i} -x_i t.$$

- If we define the vector $\mathbf{w}$ corresponding to world w as follows:

$$w_i = \begin{cases} (1-t) \text{ if } w \in b_i \\ -t \text{ otherwise} \end{cases},$$

then we can define the payoff at w from $\mathbf{x}$ as just $\mathbf{x} \cdot \mathbf{w}$.

One-Way Dutch Book 2/2

- Someone with beliefs B is subject to a one-way Dutch Book relative to t if there is some stake vector $\mathbf{x}$ that leads them to a loss in every world (a sure loss)
- We call them one-way because they only involve buying bets, not selling them.

Two-Way Dutch Book 1/2

- Lockean agents (complete ones!) ought to be willing to also sell bets at t when they do not believe them.
- A package of bets is now a n-dimensional non-negative vector $\mathbf{x}$ representing the amounts bought for each proposition in p or sold for each proposition not in p.
- Let n-dimensional vector $\mathbf{w}$ be defined as follows

$$w_i = \begin{cases} 1 - t \text{ if } p_i \in B, p_i \in w \\ -t \text{ if } p_i \in B, p_i \notin w \\ -(1-t) \text{ if } p_i \notin B, p_i \in w \\ t \text{ if } p_i \notin B, p_i \notin w \end{cases} . \qquad \text{(a)}$$

- Payoff for $\mathbf{x}$ at w is $\mathbf{x} \cdot \mathbf{w}$

Two-Way Dutch Book 2/2

- An agent is subject to a two-way Dutch Book iff there is a stake vector $\mathbf{x}$ that leads to loss at every world.
- Easy to see that if you are subject to a one-way Dutch Book then you are also subject to a two-way Dutch Book, but not vice-versa.

Accuracy Arguments

- Accuracy argument for probabilism goes by scoring graded beliefs (confidences) by how well they match the truth (e.g their accuracy).
- Accuracy arguments for probabilism show that those beliefs satisfying probability axioms have useful score-theoretic properties (e.g. being undominated)

Joyce 1998, Pettigrew 2016

A framework for scoring beliefs

- Full beliefs are simpler to score than graded beliefs.
- At any world any full belief is either right or wrong.
- We'll make these assumptions about scoring:
  - For each belief b_i the belief receives a real-valued reward $r_i \geq 0$ if it is true, and incurs a cost $s_i \leq 0$ if it is false.
  - One's total score is just the sum of ones individual scores. (Additivity)
  - The ratio $\frac{r_i}{s_i}$ is constant across beliefs.[3]

See version in Easwaran and Fitelson 2015, Easwaran 2016, and Dorst 2019.

[3] Inessential simplifying assumption.

Total Score

- The total score for a set of beliefs with at world w is:

$$S(w, b) = \sum_{w \in p_i} r_i + \sum_{w \notin p_i} s_i.$$

- Rational agents will choose beliefs to maximize their expected total score.

Reframing Scoring System 1/2

- We can redescribe the scoring system in a way that makes more clear its relationship to bets.
- Instead of having a set of r_i and s_i for each proposition p_i, let $t = \frac{-s_i}{r_i - s_i}$, and define a n-dimensional (non-negative) vector $\mathbf{x}$ such that $x_i = r_i - s_i$.
- In this case $(1-t)x_i = r_i$ and $-tx_i = s_i$.
- So, we can just define the scoring system in terms of t and $\mathbf{x}$.
- We will call $\mathbf{x}$ the weight vector

Reframing Scoring System 2/2

- In this new system we can state the agents total score at a world w as simply $\mathbf{x} \cdot \mathbf{w}$
- $\mathbf{x}$ is the o-dimensional vector weight vector for the propositions $b_1 \ldots b_o$ and $\mathbf{w}$ is the vector defined earlier:

$$w_i = \begin{cases} (1-t) \text{ if } w \in b_i \\ -t \text{ otherwise} \end{cases}$$

- Of course, t and $\mathbf{x}$ are given different interpretations here than in the betting setup.

Decision-Theoretic Properties of Scores

- Scores allow us to frame belief choice (i.e. the choice of B from P) as a decision problem.
- Rational agents with graded beliefs, for example, will maximize expected score.
- If a set of beliefs gives a negative score in every world it is a sure loss.

Dominance Notions

- The choice of one set of beliefs B strictly dominates another choice of beliefs, B' if for all worlds w, $S(w, B) > S(w, B')$.
- The choice of one set of beliefs B weakly dominates another choice of beliefs, B' if for all worlds w, $S(w, B) \geq S(w, B')$ and there is some world w', $S(w', B) > S(w', B')$.
- A set of beliefs B is not weakly/strictly dominated, if there is no set B' that weakly/strictly dominates it.

Connections

connecting the three characterizations of B

- Lockeanism about belief
- Dutch Books on sets of beliefs
- Scoring dominance of sets of beliefs

One-Way Dutch Books and Sure-Loss

For any set of worlds $W = w_1 \ldots w_m$ and set of beliefs $B = b_1 \ldots b_o$, and a positive real number t the following two statements are equivalent.

- The agent holding B is subject to a one-way Dutch Book at threshold t.
- At some weight vector $\mathbf{x}$, the agent holding B will realize a sure-loss (i.e. a loss at every world) at threshold t.

Proof.

Note that $\mathbf{x} \cdot \mathbf{w}$ (o-dimensional vectors) is both the value of a book of bets (on B) with stakes $\mathbf{x}$ with threshold t and the score from holding B with scoring system with weights $\mathbf{x}$. $\square$

Dominance and Two-Way Dutch Books

The following two statements are equivalent.

- The agent holding B is subject to a two-way Dutch Book at threshold t.
- The choice of beliefs B is strictly dominated by another set of beliefs B', on the scoring system with threshold t, and some weight vector $\mathbf{x}$.

Proof.

The proof proceeds by noting that n-dimensional $\mathbf{x} \cdot \mathbf{w}$ is both the score for taking a two-way book with beliefs B, and the score for having beliefs B minus the score for having beliefs $P \backslash B$ (so if negative you should switch, if positive you should stick with B). $\square$

Connecting Lockeanism

- Given a scoring system with threshold **t**, a set of beliefs is rational relative to some credences **c** if it is almost Lockean complete with respect to **t**.
- Or as Dorst 2019 puts in his paper title: 'Lockeans Maximize Expected Utility'.
- (Converse fails given possibility of zero weights.)

Dorst 2019; Easwaran 2016

Easwaran's Negative Observation

- Suppose a belief set B is not weakly dominated on the scoring system with weights $\mathbf{x}$ and threshold t.
- This does not entail that B is (almost) Lockean compatible.
- See example in my paper or Easwaran 2016.

Characterizing Lockean Compatibility

The following three statements are equivalent:

- The agent holding B is not subject to a one-way Dutch Book at threshold t.
- There is no weight vector $\mathbf{x}$ such that the agent holding B will realize sure-loss at threshold t.
- B is Lockean compatible with threshold t.

Notes on Proof

- The previous results followed fairly straightforwardly from definitions.
- This result, by contrast, relies on fundamental results in linear algebra relating the row-space of a matrix to its columns-space (i.e. Farkas Lemma)

Characterizing Lockean Completeness

The following three statements are equivalent:

- There is no two-way Dutch Book on B
- There is no weight vector $\mathbf{x}$ on which B is strictly dominated.
- B is almost Lockean complete.

Easwaran's Conjecture

- We can get as a corollary of the last result the "main conjecture" of Easwaran 2016.
- Easwaran's conjecture focused on scoring systems in which weights are positive rather than non-negative.
- Easwaran's conjecture: If for all positive weight vectors $\mathbf{x}$, B is not weakly dominated by some B', then B is almost Lockean complete.
- Proving just requires showing that not weakly dominated at a positive weight implies not strongly dominated at a non-negative weight.

Significance

- Lockeanism has non-derivative foundations in betting and scoring.
- Scoring requirements that characterize Lockeanism are rather strong: i.e. you need to choose a set of beliefs that is not dominated on any set of weights.
- Easwaran's programme: take all-nothing beliefs as basic as response to preface paradox. In this case the scoring requirements for theorem might be well-motivated.

Further directions

- Different scoring rules: non-additive ones.
- Connection to qualitative characterization of Lockeanism for any threshold in Adams 1965, Fernando 1998.
- Connection to Wald's Complete Class Theorem.