# Pollock's Theory of Defeasible Reasoning

Jonathan Weisberg
University of Toronto

Northern Institute of Philosophy
June 18, 2010

# Outline

# Motivations

# Motivations

1. Problematic Bayesian Idealizations
   1. Computational Demands
   2. Storage Demands
   3. Logical Omniscience
2. Representing Sequential Reasoning
3. Paradoxes of Acceptance
   1. The Lottery Paradox
   2. The Lottery Paradox Paradox
   3. The Preface Paradox
4. The Epistemological Role of Non-Doxastic States
5. Dissatisfaction with Standard Non-Monotonic Logics

# Problematic Bayesian Idealizations



- Computational Demands: a Bayesian agent updates all her probabilities with each new piece of evidence.
  - Computationally demanding, often wasteful.
  - At odds with our actual reasoning.
- Storage Demands: a Bayesian agent stores a real number for each conditional belief, a "combinatorial nightmare" (Pollock 2008).
  - Suppose an agent has 300 beliefs.
  - The number of conditional probabilities of the form $p(A|B_1 \dots B_n)$ that must be stored is about $10^{90}$.
  - $10^{90} >$ the number of particles in the universe.
- Logical Omniscience: a Bayesian agent assigns probability 1 to all logical truths, but we surely can't and don't.

Pollock (2008) advertises his framework as avoiding the first two problems.

- I'm advertising it as avoiding the last.

# Representing Sequential Reasoning

A lot of our reasoning appears to be sequential, in two ways:

- Collecting reasons.
- Deploying reasons.

Bayesianism, DST, ranking theory, etc. all ignore this reality.

- As a result, they may fail to acknowledge beliefs that are justified despite not taking account of all the evidence.
- If other cognitive demands (pragmatic or epistemic) rationally interrupt a train of reasoning, you may be justified in believing the conclusions drawn so far.

# Paradoxes of Acceptance





Pollock's treatment of the paradoxes of acceptance respects the following desiderata.

- Preface: you are justified in believing the claims in your book.
- Lottery: you are not justified in believing your ticket will win.
- Conjunction: if you are justified in believing $A$ and $B$, you are justified in believing $A \& B$.

This package is very hard to come by.

# The Epistemological Role of Non-Doxastic States



On many epistemological views, non-doxastic states play a role in justifying beliefs:

- Perceptual states
- Memories
- Module outputs

On some views, non-doxastic states alone justify:

- Pollock, Pryor

On others, they do so in conjunction with background beliefs:

- Vogel, White?

But formal epistemologies almost never address the justificatory role of non-doxastic states.

# Dissatisfaction with Standard Non-Monotonic Logics





Pollock's reasons for dissatifaction with other non-monotonic formalisms vary from case to case:

- Too limited
- Implausible results
- Off-topic

For a survey, see (Pollock 1995: 104-9).

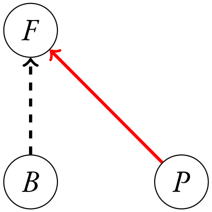
# Inference Graphs

# Epistemic States

In Pollock's system, an agent's epistemic state is represented by an inference graph.

- Nodes: reasons and the propositions they bear on.
- Directed edges: relations of support and defeat.

Example:

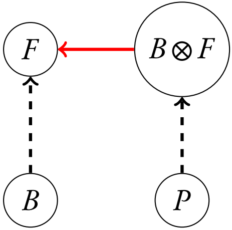


# Defeat: Rebutting vs. Undercutting

Pollock acknowledges two kinds of defeaters:

1. Rebutters: $R$ is a rebutting defeater of $P$ if it is a reason for $\neg P$.
2. Undercutters: $U$ is an undercutting defeater of $P$ as a reason for $Q$ if it is a reason for $\neg$($P$ wouldn't be true unless $Q$ were true).
   - The negated conditional is symbolized $P \otimes Q$.

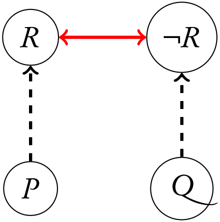
So the previous example is properly represented:



# Example: Rebutting Defeat



Example: Pam says that Robert will be at the party, whereas Qbert says he won't be:



# Inference Rules

Where do the arrows come from? That is, when is one thing a reason for another?

- Pollock proposes a number of inference rules in various writings, but does not pretend to have a complete list.
- The methodology: propose rules that seem plausible and test them on numerous examples.
  - Finding a list of complete rules that yield sensible results is a major burden of the theory.
  - Compare the Bayesian's task of specifying rationality constraints on priors: Reflection, PP, Indifference, etc.

# Inference Rules: Some Examples[1]



**Perceptual Justification** $x$'s appearing $R$ is a defeasible reason for believing that $x$ is $R$. (Pollock 1971, 1974)

**Temporal Projection** Believing $P@t$ is a defeasible reason for believing $P@(t+\Delta t)$, the strength of the reason being a monotonic decreasing function of $\Delta t$ (for appropriate $P$). (Pollock 2008)

**Discontinuity Defeat** $\neg P@t_1$ is an undercutting defeater for the inference by Temporal Projection from $P@t_0$ to $P@t_2$. (ibid)

**Statistical Syllogism** If $r > 1/2$ then $Fc \,\&\, p(G|F) \geq r$ is a prima facie reason for $Gc$, the strength of the reason being a monotonic increasing function of $r$. (Pollock 1990,1995)

**Subproperty Defeat** $Hc \,\&\, p(G|F\&H) \neq p(G|F)$ is an undercutting defeater for the Statistical Syllogism.

[1] NB: these are simplified glosses, omitting important qualifiers and details.

# Initial Nodes: Perception



Inference rules tell us how to introduce new nodes into the graph given initial nodes, but where do initial nodes come from?

- In other words, what can be used as a reason without appeal to a supporting reason?

Pollock is surprisingly brief on this point.

- Formally, we just help ourselves to a set of premises: *input*.
- He does say,

> *"Epistemic reasoning starts with premises that are input to the reasoner. In human beings, these are provided by perception." (Pollock 1995: 39)*

> *"Perception provides the premises in input from which epistemic cognition reasons forward [...]" (ibid: 47)*

# Initial Nodes: Further Candidates

Should other things be included in *input* too?

- Existing (justified) beliefs
- Memory states
- Outputs of non-perceptual modules

Fortunately, we can explore the formalism and many of its applications without answering this question.

- But it does raise important, tricky questions about what an inference graph is supposed to represent.
  - An agent's epistemic state at a time: the reasons and inferences she is currently aware of?
  - A record of her reasoning over time: all the reasons and inferences she has taken account of in her lifetime?
- The framework's appeal may depend heavily on our choice here.

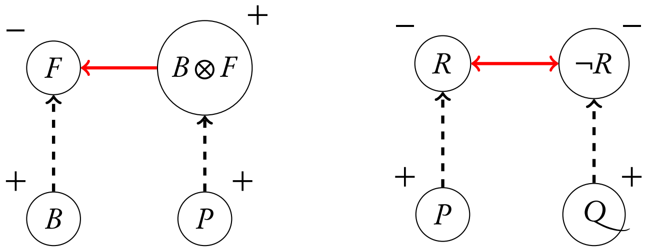
# Pollock's Semantics

# Defeat Statuses

We want to be able to figure out what beliefs are justified given the reasons and inferences taken into account so far.

- We want an algorithm for assigning the statuses *defeated* and *undefeated* to nodes in a given graph.

Using $-$ to symbolize *defeated* and $+$ to symbolize *undefeated*, we want results like:



# Semantics: A First Attempt





**Definition: D-initial Node**

A node is D-initial iff neither it nor any of its ancestors are termini of a defeat link.

Then here's a plausible, first attempt:

(1) D-initial nodes are undefeated.
(2) If the immediate ancestors of node $A$ are undefeated, and all nodes defeating it are defeated, then $A$ is undefeated.
(3) If $A$ has a defeated immediate ancestor, or there is an undefeated node that defeats $A$, then $A$ is defeated.

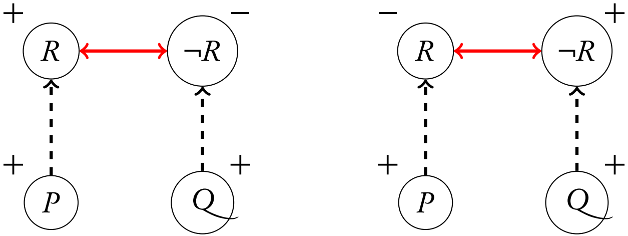
This proposal gets the right results for Tweety and other simple examples.

# A Problem: Collective Defeat



But it does poorly in cases of “collective defeat”, like our example of conflicting testimony.

- The only assignments consistent with (1)–(3) are:

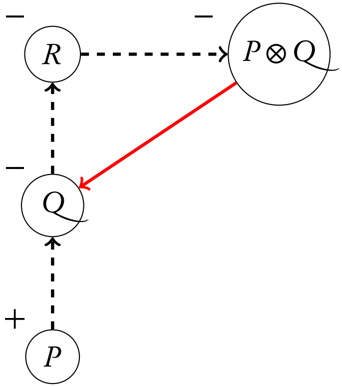



- Both are counterintuitive and unjustifiably anti-symmetric.

# Another Problem: Self-Defeat



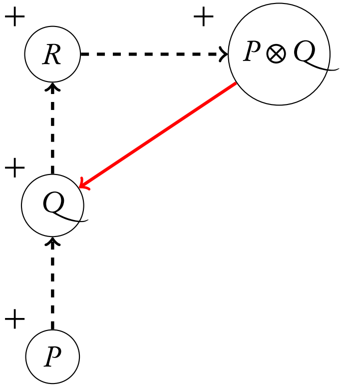
If we assign $-$ to $Q$, we violate (2):



# Another Problem: Self-Defeat



If we assign $+$ to $Q$, we violate (3):



# Partial & Maximal Assignments



These complications (and others) motivate a more sophisticated approach:

**Definition: Partial Status Assignments**

A partial status assignment assigns $+$ and $-$ to at least some nodes and satisfies:

(P1) All D-initial nodes are undefeated.

(P2) $A$ is undefeated iff the immediate ancestors of $A$ are undefeated, and all nodes defeating $A$ are defeated.

(P3) $A$ is defeated iff $A$ has a defeated immediate ancestor, or there is an undefeated node that defeats $A$.

**Definition: Maximal Status Assignment**

A status assignment is maximal iff it is partial and is not contained in any larger partial assignment.

# The Final Proposal





**Proposal: Supervaluation**

A node is undefeated iff every maximal status assignment gives it a $+$; otherwise it is defeated.

We can quickly verify that this solves our earlier problems:

- Collective Defeat: there are two maximal assignments, and $R$ and $\neg R$ each get $-$ in one of them. So both are defeated.
- Self-Defeat: there is only one maximal assignment, which merely assigns $+$ to $P$. So everything else comes out defeated.

# The Paradoxes of Acceptance

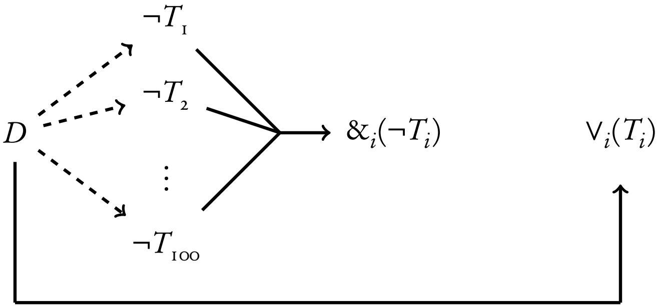
# The Lottery Paradox



A fair lottery of 100 tickets, with exactly one winner. Let

- $D$ = The description of the lottery.
- $T_i$ = Ticket #$i$ will win.

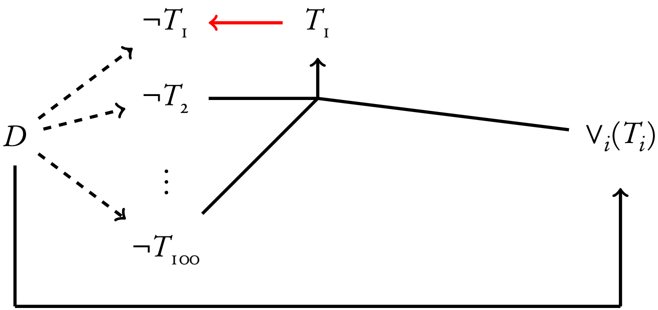
Then the paradoxical inference graph is:



# Solving the Lottery Paradox

The solution lies in noticing that there is a rebutting defeater for each $\neg T_i$.

- For example, the rebutting defeater for $\neg T_1$ is the argument for $T_1$ based on $\vee_i(T_i)$ and $\neg T_2, \ldots, \neg T_{100}$.

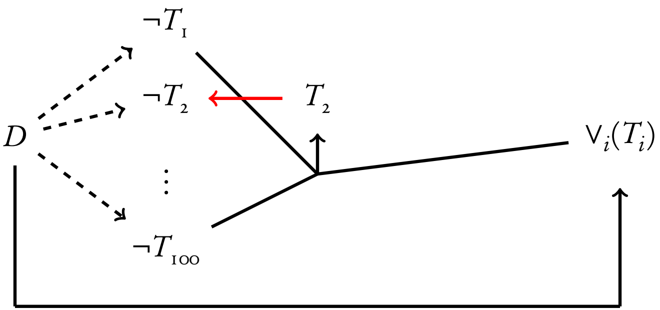


# Solving the Lottery Paradox



The solution lies in noticing that there is a rebutting defeater for each $\neg T_i$.

- Similarly, the rebutting defeater for $\neg T_2$ is the argument for $T_2$ based on $\vee_i(T_i)$ and $\neg T_1, T_3, \ldots, \neg T_{100}$.



# Solving the Lottery Paradox





Every $\neg T_i$ gets a $-$ on at least one maximal status assignment:

- For every $\neg T_k$ there is a status assignment that assigns $+$ to all the other $\neg T_i$'s and to $\vee_i(T_i)$.
- On that status assignment, $T_k$ gets a $+$.
- So $\neg T_k$ gets a $-$.

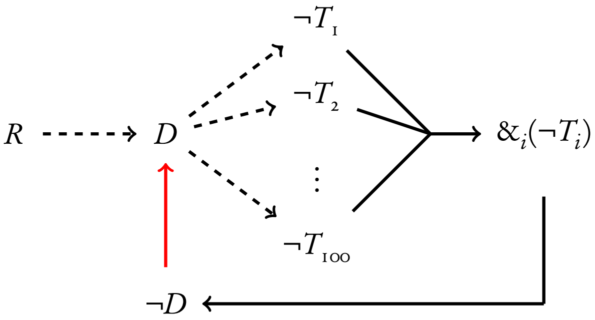
So, in the final reckoning, each $\neg T_i$ comes out defeated.

- So you are not justified in believing of any ticket that it will lose.

# The Lottery Paradox Paradox



Suppose you read about the lottery in the newspaper (R). We then have a different paradoxical challenge:




The argument has a self-defeating structure!

- So aren't we unjustified in believing the lottery will happen as described?

# Solving the Lottery Paradox Paradox



This paradox is avoided because the argument for $\neg D$ will always depend on a defeated premise.

- On every assignment, one of the $\neg T_i$ gets a $-$.
- So the argument for $\&_i(T_i)$ has a defeated premise on every assignment.
- So $\neg D$ gets $-$ on every assignment.

Pollock (2008) advertises this result as a superiority of his system over McCarthy's (1980) circumscription semantics for non-monotonic logic (and various sophistications of it).

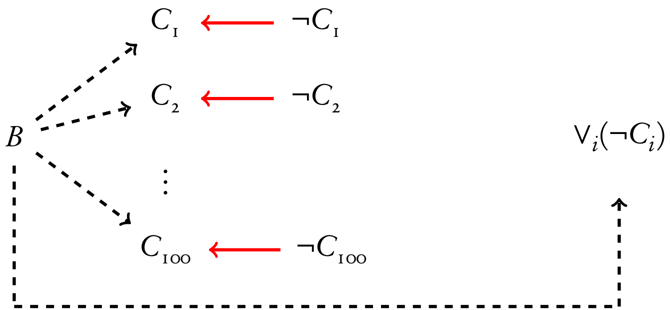
# The Preface Paradox



The preface paradox appears to have the same structure as the lottery, and so threatens to get the same, skeptical result. Let

$B$ = your background knowledge.

$C_i$ = Claim #$i$ in the book is true.



# Solving the Preface Paradox



Pollock's solution is to undermine the argument for each $\neg C_i$.

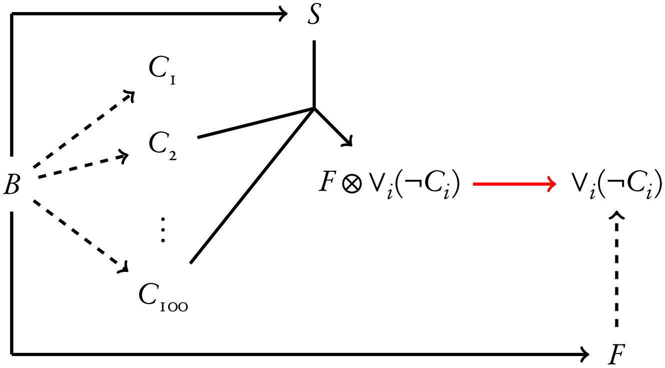
- Each $\neg C_i$ is supported by a deductive argument from the remaining $C_i$ and $\vee_i(\neg C_i)$.
- For example, $\neg C_{100}$ is supported by a deductive argument from $C_1, \ldots, C_{99}$ and $\vee_i(\neg C_i)$
- But given $C_1, \ldots, C_{99}$, the argument supporting $\vee_i(\neg C_i)$ is defeated!
    - Why? Because if the first 99 claims are true, we no longer have reason to believe that the book contains a falsehood.
    - Our reason to believe the book contains a falsehood is statistical; books of this length typically contain falsehoods.
    - But books of this length where the first 99 claims are true do not typically contain falsehoods!

# Solving the Preface Paradox

The statistical inference from $B$ to to $\vee_i(\neg C_i)$ suffers subproperty defeat on every assignment. Let

$F$: $p(\text{Falsehood}|\text{Length}) \approx 1$.

$S$: $p(\text{Falsehood}|\text{Length} \ \& \ C_2 - C_{100} \text{ are true}) \not\approx 1$.



# The Lottery vs. The Preface



Pollock's treatment of the lottery and the preface trades on a crucial difference:

- In the lottery, the $\neg T_i$ are negatively relevant to one another.
- In the preface, the $C_i$ are not negatively relevant to one another; they are either independent or positively relevant.

# Worries

# The Generalized Lottery Paradox



A threat: any proposition can be viewed as a "lottery proposition". (Korb 1992; Douven & Williamson 2006)

- Every proposition is a member of an inconsistent set of equally, statistically supported propositions.
- Thus every proposition is subject to collective defeat.

Take any proposition $P$ and a fair, 100-ticket lottery:

- Consider the set of propositions

$$\{P, \neg(P \,\&\, T_1), \ldots, \neg(P \,\&\, T_{100})\}$$

- Each member is highly probable.
- The set is inconsistent.
- So the members suffer collective defeat; none is justified.

# A Reply



Pollock explicitly qualifies the Statistical Syllogism with a projectability constraint:

- To infer that $Gc$ from the fact that $p(G|F) > r$, $G$ must be projectable with respect to $F$.
- This restriction is designed to prevent projection based on gruesome statistics.

Arguably, one's statistical evidence for a proposition like $\neg(P \,\&\, T_i)$ (if we even have such evidence) is gruesome.

- So Pollock might reply that these propositions can't even be introduced into the inference graph by appeal to SS.[2]

[2]Cf. footnote 5 of (Douven & Williamson 2006).

# Bootstrapping

Pollock's treatment of the preface threatens to lead to bootstrapping.

- Pollock is deeply committed to the Conjunction Principle.
- So you're not only justified in believing each claim in your book, you're justified in believing their conjunction!

Such immodesty has a way of fuelling itself:

- Struck by your accomplishment, you increase your estimation of your reliability as a researcher.
- Heartened, you sit down to write another book, which again turns out to be error-free!
- Lather, rinse, repeat.
- You conclude that you are infallible.

# A Shameless Plug



This problem for Pollock supports a general view I like.

- The received view: bootstrapping is a problem for *basic knowledge* theories like reliabilism and dogmatism. (Vogel 2000, 2008; Cohen 2002; van Cleve 2003)
- My view: bootstrapping is not a symptom of basic knowledge, it is a problem for everyone.
  - Bootstrapping puzzles show that justified beliefs/knowledge cannot always be used as premises in further reasoning. (Weisberg, forthcoming)

Another example: Williamson's $E = K$ thesis.

- Suppose Starla reads the first sentence in today's paper, $P$, coming to know that $P$ and that the newspaper says $P$.
- She conditionalizes her evidential probabilities on this new knowledge, increases the probability that the newspaper is reliable.

# Mixed Lotteries



Lasonen-Aarnio (2010) objects that Pollock's theory must treat "mixed" lotteries like the preface paradox:

- A mixed lottery: take one ticket from the Ontario lottery, one from the Quebec lottery, one from the Texas lottery, one from the UK lottery, etc.
- The probability of each ticket losing is very high.
- The probability of at least one winning is very high.
- But the $\neg T_i$ are not negatively relevant; they are probabilistically independent.
- So the sub-property defeat that yielded the non-skeptical result in the preface paradox should happen here too.

In short: mixed lotteries have the probabilistic structure of a preface case, so they should get the same, non-skeptical result.

# My Response



It's not clear to me that Pollock is committed to treating the mixed lottery the same as a preface.

- In a mixed lottery, each ticket is still a member of a regular lottery.
- So each $\neg T_i$ still suffers collective defeat.

In terms of defeat statuses: it is still the case that for each $\neg T_i$, there is a status assignment that gives it a $-$.

- Adding to a standard lottery graph the extra structure that comes with a mixed lottery does not rule out the status assignment that assigned $-$ to $\neg T_i$.

# Other Topics

# Variable Degrees of Justification

A natural next step is to ask how to compute defeat statuses when the degrees of justification of various arrows varies.

- See (Pollock 2001) for the details, or the expanded version online (have your LISP compiler handy).

Some notable features of Pollock's views here:

- The Weakest Link Principle: the degree of support for a conclusion of an argument is the lowest degree of support in its ancestry.
- Non-Accrual of Reasons: having more than one reason for a conclusion does not increase its degree of justification.

# Interest-Driven Reasoning





One of the most striking features of Pollock's implementation of his system for defeasible reasoning (OSCAR) is the fact that it is interest-driven.

- OSCAR doesn't just churn out theorems in some random or lexicographic order.
- It searches for answers relevant to the questions or practical problems at hand.
- The architecture for this behaviour is laid out in Chapter 4 of *Cognitive Carpentry*.

# Decisions & Planning



Pollock (1995: 179-183) rejects standard decision theory.

- Standard decision theory overlooks the importance of planning.
- The Button Problem: if you press buttons $A$, $B$, $C$, and $D$, you get £10; if you press button $E$ you get £5.
- Pollock argues that, on standard decision theory, pushing button $A$ does not maximize expected utility.

Pollock (1995: ch. 5) opts for a two-tier theory of practical reasoning:

- Agents fist construct plans aimed at goals.
- They then choose plans based on expected utility maximization.

# References I



[1] Stewart Cohen.
Basic knowledge and the problem of easy knowledge.
*Philosophy and Phenomenological Research*, 65(2):309–329, 2002.

[2] Maria Lasonen-Aarnio.
Is there a viable account of well-founded belief.
*Erkenntnis*, 72(2):205–231, 2010.

[3] John L. Pollock.
Perceptual knowledge.
*The Philosophical Review*, 80(2):287–319, 1971.

[4] John L. Pollock.
*Knowledge and Justification*.
Princeton University Press, 1974.

[5] John L. Pollock.
The paradox of the preface.
*Philosophy of Science*, 53(2):246–258, 1986.

[6] John L. Pollock.
*Cognitive Carpentry: A Blueprint for How to Build a Person*.
MIT Press, 1995.

[7] John L. Pollock.
Defeasible reasoning with variable degrees of justification.
*Artificial Intelligence*, 133(2):233–282, 2001.

[8] John L. Pollock.
Defeasible reasoning.
In Jonathan E. Adler and Lance J. Rips, editors, *Reasoning: Studies of Human Inference and its Foundations*.
Cambridge University Press, 2008.

# References II



[9] James Van Cleve.
Is knowledge easy — or impossible? externalism as the only alternative to skepticism.
In *The Skeptics: Contemporary Essays*. Aldershot, UK: Ashgate Publishing, 2003.

[10] Jonathan Vogel.
Reliabilism leveled.
*Journal of Philosophy*, XCVII(11):602–623, 2000.

[11] Jonathan Vogel.
Epistemic bootstrapping.
*Journal of Philosophy*, CV(9):518–539, 2008.

[12] Jonathan Weisberg.
Bootstrapping in general.
*Philosophy and Phenomenological Research*, forthcoming.