# The Myth of a Confirmation Bias (and How to Get Rid of It)

## Arguments for a Better Argumentative Theory of Reasoning

Emmanuel J. Genot (Lund, Theoretical Philosophy)

April 11, 2013

# Introduction

# Confirmation Bias: There and Back Again



- Experiments carried in the 1960s showed that subjects testing a conditional rule "**If P then Q**" tend to infer **P** from **Q**—as per the **Fallacy of Affirming the Consequent**.
- Initial explanation: subjects try to **confirm that the rule holds**—incorrect according to Popper's Fallibilism, popular in psychology at the time—hence the hypothesis of a **Confirmation Bias** (**CB**).
- In the 1990s, **Bayesian psychologists** (opposed to Fallibilism) argued that the problems submitted were **not deductive** and that **subjects' choices were in line with a statistical solution**—no fallacy, no **CB**.
- In the 2000s, **Relevance Theorists** used new experiments to argue that subjects' response is **too fast for reasoning** (deductive or statistical), guided instead by **'intuitions of relevance'** (not much reasoning, but no fallacy).
- The **Argumentative Theory of Reasoning (ATR)** took over: intuitions select **'self-serving' argumentative strategies** internalized by natural selection—at cross-purpose with 'deductive competence'—**CB is back with a vengeance**!

# The Argument



1. The hypothesis of **CB** depends on the assumption that humans have a **deductive competence** that enables **classical logical reasoning** if unhindered.

2. The hypothesis unduly **privileges one logic** (classical) that applies to a **limited class of contexts**, that do not exhaust natural language uses.

3. **Bayesian and Relevance-theoretic explanations are 'deductive'**—when 'deduction' is understood relative to the **appropriate semantics**.

4. **External goals may be necessary** to differentiate between semantically equivalent solutions (some may be 'argumentative').

5. There is more empirical evidence that **self-serving argumentative agendas do not always obfuscate the truth** than to the contrary.

# Will Do



1. Present **Wason's Selection Task**—the experiment that started it all—with its **three 'worked out' explanations**:
   (a) Wason's **'deductive'** explanation—with fallacy/**CB**;
   (b) the Bayesian **'inductive'** explanation—without fallacy/**CB**;
   (c) Relevance Theory **'pragmatic'** explanation—without reasoning (without fallacy/**CB**).

2. Present a semantics **capturing subjects interpretation of conditionals** that:
   (a) **generalizes Bayesian models** with 'looser' semantic constraints;
   (b) partially **explicates 'pragmatic' intuitions** semantically;
   (c) leaves room for **contextual goals** to be 'factored in'.

3. Conclude that an 'argumentative' account can **improve on current theories**, and in particular:
   (a) **unify them**: 'rational analysis of intuitions of relevance';
   (b) **crowd out CB** and other myths inherited from inapropriate semantics;
   (c) more generally **update the research agenda** in cognitive psychology.

# The Selection Task(s)

# "The Mother of All Reasoning Tasks"



Below is a set of **four cards**, with a number on one side and a letter on the other; and a **rule** that applies only to them. Which (if any) of these four cards **must be turned** in order to decide if the rule is true, without turning unnecessary cards?

**If there is a vowel on one side,**

**there is an even number on the other side** (Rule)

# Classical Deductive Solutions



- (Rule) simplifies in **If P, then Q**: only **propositional reasoning** is needed (but subjects are often puzzled by quantification [SvL08]).
- Assuming a **material conditional**, classical schemas of **Modus Ponens (MP)** and **Modus Tollens (MT)** apply:
  - **(MP)** with A **predicts** 4 on the hidden side;
  - **(MT)** with 7 **predicts** K on the hidden side;
- If **both** predictions are confirmed, (Rule) holds; if **either** fails, (Rule) does not hold (no need to try for 'falsification'!).
- The **deductive solutions** with material conditional combine the '**P card**' A and the '**not-Q card**' 7 (Wason and others assumed mistakenly a unique solution [Gen12]).
- These solutions **cannot account for the data:**
  A alone: 35%; A, 4: 45%; A, 7: 5%; A, 4 and 7: 7%; misc.: 8%

# Bayes or Bust : The Rational Analysis of the Selection Task (I)



*"The selection task is a 'loose' probabilistic task, rather than a 'tight' deductive task."* [OC94, p. 626]

The **Rational Analysis of the Selection Task (RAST)** recovers from the instructions in ST an **initial probability to be in a** (Rule)-**state**, and an **update rule**.

**Initial Probability** **(a)** there are less vowels than consonants than even numbers than numbers; **(b)** cards are a **sample** from a population where no letter/number pair is ruled out; **(c)** with no other knowledge about the population of cards, proportions should be assumed as per **(a)**.

**Updating 1** Revealing values at the back of $(X, \cdot)$ incur a **revision of the initial probability to be in a** (Rule)-**state**—'farther away' or 'closer by' than hypothesized based on **(a)**;

**Updating 2** Cards can be compared w.r.t. the **amplitude of the variation** they can cause.

# Bayes or Bust : The Rational Analysis of the Selection Task (II)



- With **Initial Probability** and **Updating** so defined, there is a **unique preference ordering for data selection** $\succ$:
  $(A, \cdot) \succ (4, \cdot) \succ (7, \cdot) \succ (K, \cdot)$—or:

  $$\textbf{P} \succ \textbf{Q} \succ \textbf{not-Q} \succ \textbf{not-P}$$

- Subjects behave **'as if' performing optimal data selection** in a statistical problem **under assumptions (a)-(c)**.
- Assuming **"Bayesian Brain" hypothesis**—that statistical reasoning has been selected throughout evolution—together with **assumptions (a)-(c)**, there is **no deduction and no bias**.

# Inept Reasoners vs. Pragmatic Virtuosos (I)



*"Regarding the common view that selection task results show that people do not reason in accordance with the rules of logic, the fluctuating results we obtained [ suggest] either that people are even worse reasoners than was claimed, or, more plausibly, that they are pragmatic virtuosos.* [GKSvdH01, B75]

**Relevance Theorists** propose that selection respond to **contextual cues**, pertaining to the role of information in communication. Their experiments in [GKSvdH01] use:

- A unique rule:
  **If a person travels to any East African country, then that person must be immunized against cholera**.
- Different narratives where the subject is **a travel agent** and:

  **Task 1** at first, **believes that the rule is in force**, and tries to convince a client who believes otherwise;

  **Task 2** later, **believes that the regulation has changed**, and tries to convince her boss who believes otherwise.

# Inept Reasoners vs. Pragmatic Virtuosos (II)



- Relevance Theory contends that "the very process of linguistic comprehension of the problem yields **intuitions of relevance** that, in turn, determine card selection" [GKSvdH01, B75].
- Their predictions are accordingly 'intuitive':

**Task 1** *"the rule* **achieves relevance to the audience** *(i.e. to the client of the agency) [...] by implying that any given East African country would require immunization, and that among countries requiring immunization are East African countries. We predicted [that subjects] would choose the* **P** *(an East African country) and the* **Q** *(a country requiring cholera immunization) cards"* [GKSvdH01, B73]

**Selections:** **P,Q**: 69%; $P$, **P,not-Q**: 9%; **P,Q**, **not-Q**: 9%; Other: 17%

**Task 2** *" The boss's assertion* **achieves relevance** *by implying that any given East African country would require immunization and that there is no East African country that does not require it. We predicted [that subjects] would select the* **P** *(an East African country) and the* **not-Q** *(a country not requiring cholera immunization) cards."* [GKSvdH01, B73]

**Selections:** **P,Q**: 15%; **P,not-Q**: 47%; **P,Q, not-Q**: 24%; Other: 14%

# Rather than Intuitions or Probabilities

# Semantics for Conditionals (I): A Tight One



- **Adams' semantics generalizes classical semantics** for propositional logic; in particular:
  - Propositions are assigned **real values** in the interval $[0, 1]$, rather natural numbers 0 or 1 only (constraints on probability distributions apply).
  - Classical inferences are **valid for extreme values** (0 and 1) but not always otherwise
- Spontaneous inferences from **natural language conditionals** agree with inferences warranted by **Adams' logic**—which encodes the rules valid according to Adams' semantics [GO12].
- When initial probabilities as fixed for ST with **(a)-(c)**, subjects selection in ST is **deductive according to Adams' semantics** [Duc09].

# Semantics for Conditionals (II): A Loose One



- When the precise probabilistic metric can be neglected, Adams' truth-conditions for conditionals "**If P then Q**" can be relaxed to:

  Given **P**, **Q** is strictly more likely than **not-Q** (**Cond**)

- The 'loose' (**Cond**) **can be tightened** in:
  - **Adams' semantics** (aka probabilities) with specific assumptions (letters, numbers, countries, immunizations, etc.);
  - **Classical semantics** if "more likely" is arbitrarily high, no **not-Q** is admissible.

- **Received interpretations are too 'tight' for ST**: Wason's is incorrect, and a 'loose' (**Cond**) is sufficient to for RAST.

- Loosening Adams' conditional to (**Cond**) can also **explain semantically 'intuitions of relevance'** (partially, but more to come).

# A Four-card poll



- Within a sample truth-conditions (**Cond**) are satisfied iff '**P and not-Q**' cases **do not outnumber** '**P and Q**' cases.
- **Relevance** for assessing (Rule) depends on whether paired values count as a *pro*, *con*, or neither:

| **Relevance** | Cards (1) | Cards (2) |
|---|---|---|
| **Pro** | $(A,4)$, $(4,A)$ | $(P,Q)$, $(Q,P)$ |
| **Con** | $(A,7)$, $(7,A)$ | $(P,\overline{Q})$, $(\overline{Q},P)$ |
| **Neutral** | $(4,K)$, $(7,K)$, $(K,4)$, $(K,7)$ | $(Q,\overline{P})$, $(\overline{Q},\overline{P})$, $(\overline{P},Q)$, $(\overline{P},\overline{Q})$ |

- The **P-card** is **unconditionally relevant**; the **not-P** unconditionally **irrelevant**; other cards are **conditionally relevant** (by exclusions and dependencies).
- Flipping all **possibly relevant cards** is **P, Q, not-Q** is not always necessary, and subjects use **contingency plans** [SvL08]:

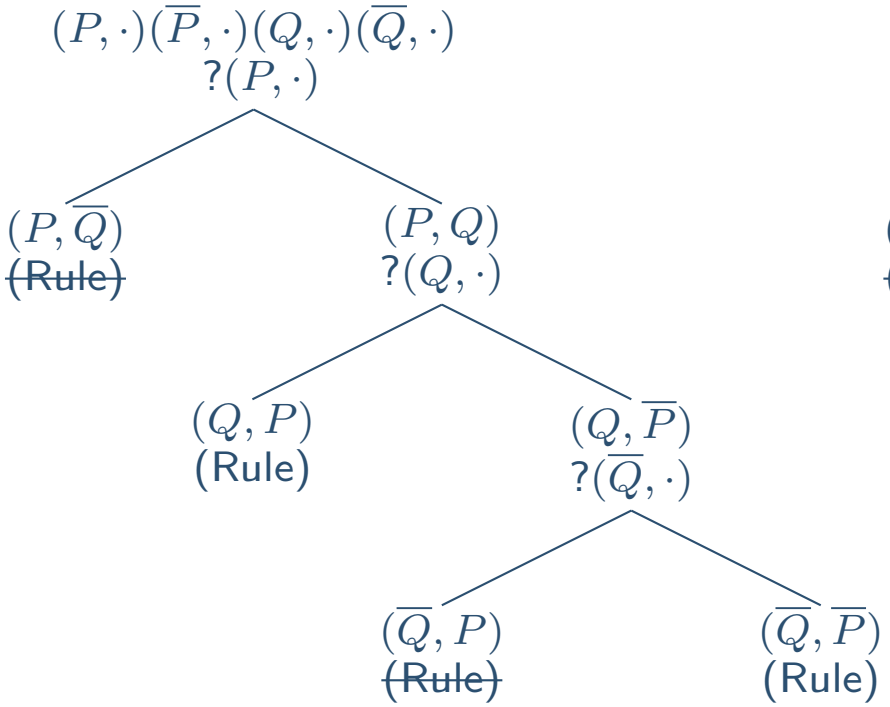
  **Pb 1** (Subject) Which is the **best selection to report** (shortest, longest...)?

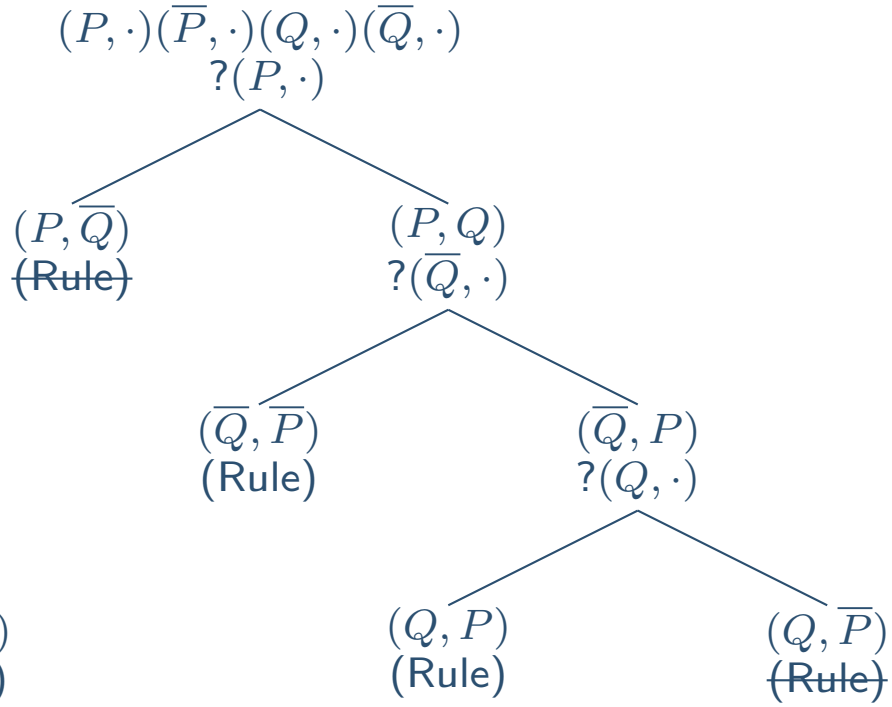
  **Pb 2** (Experimenter) How to **infer the strategy from the report**? (No good solution.)

# Solution Paths with (Cond)



Considering **contingency plans** with (**Cond**), there are **two optimal solution paths**:




**Note:** Each 'leaf' (terminal node) is an **information fixpoint** where one knows whether there is a majority count for the rule—not always what the majority is .

# Symmetry (and How to Break It)



- (**Cond**) **does not favor any one solution path**: any mixture of the two is acceptable, with the following order $\succ^*$.

$$\mathbf{P} \succ^* \mathbf{Q} \sim^* \mathbf{not\text{-}Q} \succ^* \mathbf{not\text{-}P}$$

- Strengthening **semantically** $\succ^*$ is possible—with a 'tighter' semantics, as e.g. with the Bayesian $\succ$—but this is not necessary.
- Alternatively $\succ^*$ can be strengthened by aggregation of a **secondary preference orderings**—in [GKSvdH01] **argumentative positions** seem to play such a role (possibly also to choose a **report**).
- No room for either biases or 'classical' deductive competence:
  - material conditional is 'too tight'; **classical semantics does not apply**.
  - tie-breaking **preferences compatible with data** discriminate between equally good **truth-tracking selection histories**

# Conclusions

# Summary



- **Logical validity** is always relative to a given semantics, and **valid inference schemas are relative a given semantics** too:
  - Classical schemas (e.g. **Modus Ponens** and **Modus Tollens**) are valid relative to **classical semantics**, but not always relative to all others;
  - Conversely, **classically invalid schemas**—like "Affirming the consequent"—can be **admissible with semantics other than classical**.
- Experimental subjects understand **natural language conditionals** not as material conditionals but as **statements of 'relative expectedness'** [GO12]:
  - **Bayesian models** that account for this phenomenon are equivalent to 'Adams' semantics' which is a **generalization of classical semantics**;
  - Statistical solutions in rule-testing tasks are **deductive relative to that semantics** [Duc09].
- **Probabilities are unnecessary**, and its good news:
  - **sample-bounded reasoning** differing projection has dramatically lower computational cost in ST (important for models of real-life reasoners [BS92]);
  - **external argumentative goals** suffice as **tie-breakers** (no need to tighten the semantics). **[Work in Progress]**

# An Alternative Agenda



Theories resting on 'deductive competence' and 'biases' (including **ATR**) comply with the **research agenda for psychology of reasoning** set 10 yrs ago (cf. [Eva02]).

**Local** For ST and other specific tasks: identify **relevant semantic interpretations**, including:

1. possible **mismatch b/w Subjects and Experimenters**;
2. **'pragmatic' tie-breakers** for semantically equally good strategies;
3. **specific algoritmic costs** for planning strategies and reports.

**General** (for data on reasoning) Propose a **3-tier model of experiments** that covers:

1. **Coordination on instructions** and their interpretation;
2. **Planning of strategies** including the role of contingencies;
3. **Selection of reports** including cost-benefit of messages.

# Acknowledgements



Thanks to Can Bashkent, Catarina Dutilh Novaes, Sebastian Enqvist, Peter Gärdenfors, Yacin Hamami, Justine Jacot, Erik J. Olsson, Marco Panza, Philip Pärnamets, Patricia Rich, Gabriel Sandu & Frank Zenker.

and thank you!

[BS92] Ken Binmore and Hyun Song Shin. Algorithmic knowledge and game theory. In Cristina Bicchieri and Maria Luisa Della Chiara, editors, *Knowledge, Belief and Strategic Interaction*, Cambridge Studies in Probability, Induction, and Decision Theory, chapter 9, pages 141–154. Cambridge University Press, 1992.

[Duc09] Simone Duca. Rationality and the Wason selection task: A logical account. *Psyche: An International Journal of Research on Consciousness*, 15(1):109–131, 2009.

[Eva02] Jonathan St.B.T. Evans. Logic and Human Reasoning: An Assessment of the Dedcutive Paradigm. *Psychological Bulletin*, 126(6):978–996, November 2002.

[Gen12] Emmanuel J. Genot. Information Processing and Constraint Satisfaction in Wason's Selection Task. In Jesus M. Larrazabal, editor, *Proceedings of the ILCLI International Workshop on Cognitive Science*, Donostia - San Sebastian, 28–30 November 2012. University of the Basque Country Press.

[GKSvdH01] Vittorio Girotto, Markus Kemmelmeier, Dan Sperber, and Jean-Baptiste van der Henst. Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition*, 81(2):B69–B76, 2001.

[GO12] Angelo Gilio and David Over. The psychology of inferring conditionals from disjunctions: A probabilistic study. *Journal of Mathematical Psychology*, 56(2):118–131, 2012.

[OC94] Mike Oaksford and Nick Chater. A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101(4):608–631, 1994.

[SvL08] Keith Stenning and Michiel van Lambalgen. *Human Reasoning and Cognitive Science*. The MIT Press, Cambridge (Mass.), 2008.