Information Processing and Constraint Satisfaction in Wason’s Selection Task

Genot, Emmanuel


2012

Link to publication


Total number of authors: 1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Information Processing and Constraint Satisfaction in Wason’s Selection Task∗

Emmanuel J. Genot
Department of Philosophy, University of Lund
emmanuel.genot@fil.lu.se

October 12, 2012

Abstract

In Wason’s Selection Task, subjects: (i) process information from the instructions and build a mental representation of the problem, then: (ii) select a course of action to solve the problem, under the constraints imposed by the instructions. We analyze both aspects as part of a constraint satisfaction problem without assuming Wason’s ‘logical’ solution to be the correct one. We show that outcome of step (i) may induce mutually inconsistent constraints, causing subjects to select at step (ii) solutions that violate some of them. Our analysis explains why inconsistent constraints are less likely disrupt non-abstract (or “thematic”) versions of the tasks, but unlike Bayesians does not posit different mechanisms in abstract and thematic variants. We then assess the logicality of the task, and conclude on cognitive tasks as coordination problems.

1 Introduction

In Wason’s Selection Task (hereafter ST, see e.g. [20]) subjects must test a conditional rule over a four-card setting. Empirical results do not confirm ‘logical’ predictions, and have initially been interpreted as evidence of systematic cognitive biases. Although subjects perform better when the content of the rule is ‘thematic’ content, i.e bears upon familiar social contexts, their success is attributed to social skills (like ‘cheater detection’). In the 1990s, Bayesians proposed that subjects in fact address abstract rules statistically, and ‘thematic’ ones deontically [13]. Average responses in the former matches optimal data selection over a sample. In the latter, a strong preference for upholding the rule exhaust violators, and emulates a ‘logical’ selection. This model was challenged in the early 2000s by Relevance theorists, who manipulated performance independently of the content of the rule, and concluded that little reasoning (deductive or otherwise) actually occurs in ST[5]. Nevertheless, Bayesians and Relevance theorists alike accept that if deductive reasoning was carried in ST, Wason’s ‘logical’ solution should be implemented.

K. Stenning and M. van Lambalgen have proposed a more sophisticated semantic analysis of ST, according to which solving ST (in any of its variants) is a two-step process, where subjects first process information from the instructions, and recover a representation of the problem, then plan and execute a course of action to solve it [16–18]. Both steps can be carried with varying degrees of awareness, and can e.g. involve ‘precomputations’ of different origins. While they have extensively studied how ambiguities of instructions affect the first step [18, ch. 3], their have given less attention to the second, nor attempted to reconstruct

∗This paper is based on a joint work with Justine Jacot.
the task and assess its ‘logicality.’ This paper fills this gap, and examines the extend to which ST is a ‘logical’ problem, independently on whether the problem-solving engine is a reasoner, or an abstract computer. This approach nonetheless illuminates the empirical task (but we contain its discussion to footnotes and conclusion, for space reasons).

Section 2 reconstructs Wason’s interpretation of ST and his favored solution as resp. a set of constraints and a claim about their optimal satisfaction, then introduces a formal model in order to evaluate that claim. Section 3 translates ST in the model, shows its constraints to be jointly unsatisfiable, and discusses variants solvable as intended, pending revision of the interpretation of instructions. Section 4 discusses whether any of these problem is indeed ‘deductive’ and we conclude on relations between our model, Bayesian models, and Relevance Theory, and the issue of coordination between subjects and experimenters.

2 Background and Method

A typical formulation for the abstract version of ST (borrowed from [18, p. 44]) is as follows:

Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other. Also below there is a rule which applies only to the four cards. Your task is to decide which if any of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the cards you want to turn.

**Rule** If there is a vowel on one side, then there is an even number on the other side.

**Cards**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>4</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>K</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wason’s viewed ST as a laboratory model of scientific hypothesis testing, and initially adopted Popper’s fallibilism as a normative model [14]. Assuming that Rule is as a (material) conditional if $P$ then $Q$, falsification applies *Modus Tollens*, which warrants from Rule and not-$Q$ an inference to not-$P$. Wason took detection of potential $P$-and-not-$Q$ cases to be methodologically sound and deductively competent, yielding as normative selection both and only the $P$-card (A, ·) and the $Q$-card (7, ·) (where in $(X, Y)$, $X$ is the initially visible face first, and · an unknown value). In the above variant, however, less than 10% of subjects conform to expectations, with as typical results: (A, ·) alone 35%; (A, ·) and (4, ·), 45%; (A, ·) and (7, ·), 5%; (A, ·), (4, ·) and (7, ·), 7%; and other selections, 8% [18].

Wason’s intended reading of the instructions can be rephrased as a set of constraints over the representation of ST and its admissible solutions. Subjects must limit their attention to the four-cards setting, with 4 or 7 (resp.: A or K) as only possible back values for cards (A, ·) and (K, ·) (resp.: (4, ·) and (7, ·)) (C1). They must consider no other action than turning a subset of these cards (C2), interpret Rule as a material conditional (C3), pick a strategy that decides Rule holds and report its selection (C4), and must not turn unnecessary cards (C5).

None of (C1)–(C5) is guaranteed by the instructions. Bayesians exploit the hypothesis that (C1) is never (or almost never) retrieved with ‘abstract’ rules, but substituted with sample reading. Correlations between variation in formulations and in responses have been extensively studied, and space reasons prevent us to cover them (for a survey, see in [18, 18]).

---

1 The rule quantifies (implicitly) over sides, numbers and letters, and propositional simplification is causes difficulties of interpretation that may e.g. elicit a ‘sample’ reading [18, ch. 3].

2 Bayesians do not view hypothesis testing as a deductive task, and contend that subjects address the task with (implicit) knowledge that there are less vowels than consonant, less letters than even numbers, and less even numbers than non-even ones, inducing preference ordering statistically testing the rule as follows: $(A, \cdot) > (4, \cdot) > (7, \cdot) > (K, \cdot)$, and matching experimental data [13, p. 625].
§ 3.7], esp. tables pp. 88–89). Yet, there is a large consensus that Wason’s selection is normative under the above constraints, which translates as follows:

**Postulate A.** Under conditions (C1)–(C5), the selection including exactly the (A, ) and (7,) dominates in ST.

Other reconstructions of Wason’s normativity hypothesis are possible (cf. § 3), but evaluating any candidate requires a model of the problem characterized by (C1)–(C5). This problem can be studied within the interrogative model of inquiry (IM) propose by J. Hintikka and his associates [6, 7]. The IM builds upon an abstract model where (given some background theory T) Inquirer investigates a principal question Q, by means of instrumental questions supplementing her information about the underlying state of Nature. In the simplest case, Q partitions the possible states of Nature compatible with T, denoted hereafter SoN(T). The IM is a special case of (algorithmic) learning models, in which solvability of (T, Q) depends on the existence of specific learning strategies [10, 12].

A learning strategy σ is a function taking as argument finite segments of possibly infinite sequences of observations (where an infinite sequence can in the limit fully characterize the underlying state of Nature), and returns either an answer in Q, or ‘?’ (suspension). In the IM, σ additionally affects how the data sequence is collected. Whenever some state S ∈ SoN(T) satisfies some q_i ∈ Q (in symbol: S ⊨ q_i), σ solves ⟨T, Q⟩ in S iff for every sequence of observations characterizing S, there is a finite segment after which σ outputs q_i ∈ Q and never changes its assessment. Accordingly, σ solves ⟨T, Q⟩ simpliciter iff σ solves ⟨T, Q⟩ in every S ∈ SoN(T).

Implementing a learning strategy “does not amount to ‘waiting for deduction to work’” [12, p. 1354], even for strategies ‘tracking’ deductibility, aka patient strategies, where σ is patient whenever it outputs some q_i ∈ Q in S only if q_i is deductible from T and the information gathered about S. Indeed, σ can be the functional description of some underlying mechanisms that is not ‘logical’ in any particular sense. The model remains noncommittal about mechanisms, and need not postulate e.g. that some ‘logical competence’ backs humans’ implementation of patient strategies.

Background information and answers generates an information bipartition over SoN(T), with a cell comprising scenarios compatible with the answers, and the other, those which are not. Prior to any answer being received, the first cell is identical with SoN(T), i.e. all possible states are indiscernible,\(^3\) and this partition is refined when incoming answers ‘hack off’ scenarios compatible with them. A strategy is credulous whenever it accepts an answer as soon as obtained, and is cautious otherwise. A credulous σ never requires confirmation, but may have to retract acceptance (aka ‘bracket’ answers or parts of T, cf. [3, 7]). When all answers are true, σ never has to do so, and refines indiscernibility no slower, and sometimes faster, than any cautious σ’ identical to σ save possibly w.r.t. confirmation requests; since σ then identifies S_0 ‘up to’ inclusion in some q_i ∈ Q, no later than σ’, and possibly earlier, a credulous σ (weakly) dominates (ceteris paribus) a cautious σ’ in a problem with true answers.

When finitely many parameters suffice to identify S_0 up to inclusion in some q ∈ Q is finite, and all the values for those parameters are available in S, ⟨T, Q⟩ is decidable in S, and any method solving it can (in principle) halt on success. Decidability is a special case of solvability: undecidable problems are unsolvable by either halting or patient strategies,

\(^3\)“Indiscernible” is often paraphrased as: “indiscernible for all X knows” where X is an an agent with ‘intentional’ states (beliefs, expectations, preferences, etc.). However, it can be relativized to any test that identifies some property of a state but not others: e.g. for a thermometer, states are indiscernible w.r.t. humidity. Accordingly, a ‘solving engine’ can be an abstract or mechanical device (devoid of awareness), or an agent with intentional states (awareness remains optional).
but are sometimes solvable in the limit by ‘impatient’ ones [11]. For a decidable problem, a patient \( \sigma \) that exhausts all and only the relevant parameters always solves (and decides, if halting) a problem no later than an impatient \( \sigma' \) identical with \( \sigma \) save possibly for the initial assessment, an sometimes earlier if \( \sigma' \) must revise it, i.e. ceteris paribus a patient \( \sigma \) (weakly) dominates an impatient \( \sigma' \) in a decidable problem.

By the above arguments, patient-credulous (p-c) strategies dominate (ceteris paribus) in decidable problems with true answers. A p-c \( \sigma \) can still halt later than success. In particular, if in some \( P \) the overall set of parameters for \( S \) is finite, if \( \sigma \) exhaustively enumerates them, \( \sigma \) solves \( P \) no later (ceteris paribus) than an impatient or cautious \( \sigma' \), but may decide \( P \) later than \( \sigma' \) if \( \sigma' \) collects only sufficiently many parameters to solve \( P \) and halts on success, while \( \sigma \) is still collecting. Some constraints on strategies can sometimes optimize solution (or decision), and a constrained learning problem \( P \) is a triple \( P = (T, Q, C) \), where \( C \) is a set of constraints. Notice that, when the constraints are too strong, \( P \) may not be solvable even if \( P \) minus \( C \) is .

3 Results

The intended interpretation of ST given by (C1)–(C5) characterizes a constrained learning problem \( P_{ST} = (T_{ST}, Q_{ST}, C_{ST}) \), whose properties are independent on whether the problem solver is, human, mechanical or abstract, has awareness or not, etc. (C1) is equivalent to the part of \( T_{ST} \) that specifies (four) relevant parameters of every \( S \in \text{SoN}(T_{ST}) \) as follows:

\[
\text{SoN}(T_{ST}) = \{(A, x_1), (K, x_2), (4, x_3), (7, x_4) \} : x_1, x_2 \in \{4, 7\}, x_3, x_4 \in \{A, K\}
\]

We let \( S_0 \in \text{SoN}(T_{ST}) \) denote the underlying state of Nature (which can be any of sixteen states compatible with \( T_{ST} \)). (C1) does not constrain relative probabilities, but rules out that \( S_0 \) depends on learning strategies (selections only complete the initially incomplete information). Abusing notation, we represent initial information about \( S_0 \) as:

\[
\forall \sigma, \quad \text{Inf}^\sigma_0(S_0) = \{(A, \frac{3}{4}), (K, \frac{3}{4}), (4, \frac{1}{2}), (7, \frac{1}{2})\}\sigma
\]

where in \((X, \frac{3}{4})\), \( X \) is known, and there is uncertainty between \( y_1 \) and \( y_2 \). \( \text{Inf}^\sigma_0(\text{SoN}) \) can be expanded to a truth-table-like 16-rows matrix representing the equivalence class of states indiscernible from \( S_0 \), identical with \( \text{SoN}(T) \) prior to any answers.

(C2) tells that values of those parameters can always be obtained, if it does not rule any of the possible selections (we note a selection turn[X] with \( X \subseteq \{(A, \cdot), (K, \cdot), (4, \cdot), (7, \cdot)\} \)). Both one-shot strategies (making a unique selection before they stop) and sequential ones (making several successive selections, possibly with contingency plans) are allowed. (C2) also implies that the information about the back of cards is reliable (at least as much as a visual check is). (C3) bipartitions \( \text{SoN}(T) \), through the truth-functional meaning of if . . . then as follows:

\[
\text{Rule} = \{(A, x_1), (K, x_2), (4, x_3), (7, x_4) : x_1 = 4 \text{ and } x_4 = K\} \quad \text{(Rule)}
\]

\[
\overline{\text{Rule}} = \{(A, x_1), (K, x_2), (4, x_3), (7, x_4) : x_1 = 7 \text{ or } x_4 = A\} \quad \text{(\overline{Rule})}
\]

A paradigmatic undecidable problem is the halting problem for a program that runs either finitely or infinitely many steps, and asks for a \( \sigma \) which, after witnessing only finitely many steps, decides whether the program will run finitely or infinitely. No patient or halting strategy can solve the problem when the program does not. However, the ‘impatient’ \( \sigma \) that initially conjectures that it does not, and repeats this conjecture indefinitely unless the program stops (in which case it states it, and halt) solves the problem on every run of the program.

*Human reasoners must first solve information processing problem, i.e. recover a mental representation of \( P_{ST} \) from proximate linguistic interaction, coordinated with the experimenter’s interpretation (see § 5)*

Instructions expressing (C4) typically do not govern reports of contingency plans, and heuristically favor one-shot strategies, but nothing in principle prevents subjects to contemplate sequential ones, and tutorial experiments where subjects explain their selections show that at least some do [18, § 3.5].
and yields a test to assess $Q_{ST} = \{ \text{Rule}, \text{Rule}^c \}$, because characteristic properties of Rule and Rule$^c$ mention observable values.

In what follows, ‘test for $Q_{ST}$’ or simply ‘perform a test’ abbreviates: “test for the (observable) characteristic property of Rule or Rule$^c$ and output an answer if possible, otherwise return ‘?’”, because (C1)–(C3) and the reliability of (outputs of) selections entail that patient-credible strategies (weakly) dominate (ceteris paribus) in $P_{ST}$. No strategy can then satisfy Post. A and strictly dominate a r-c $\sigma$ that also satisfies it, and therefore we restrict our attention to only those.

The first explicit constraint is (C4) imposing decision. Since $P_{ST} \setminus C_{ST}$ is already decidable in principle, the crux is (C5), which excludes strategies using unnecessary questions, i.e. questions that do not contribute to apply the test for $Q_{ST}$. Finally, (C3) and (C4) together imply that in order to perform a test, $\text{Inf}^0_p(S_0)$ must shrink down to at least:

\[
\text{Inf}^0_p(S_0) = \begin{cases} 
(A, 4), (K, [\frac{3}{4}]), (4, [\frac{3}{4}]), (7, K) & \text{if } S_0 \in \text{Rule}; \\
(A, 7), (K, [\frac{3}{4}]), (4, [\frac{3}{4}]), (7, K) & \text{if } S_0 \in \text{Rule}^c.
\end{cases}
\]

(C5) prevents selections to include $(K, \cdot)$ or $(4, \cdot)$, which are always unnecessary. How ‘necessity’ is to be appreciated in other cases is left unspecified. However, a natural suggestion is to start with some $\sigma$, check whether it solves $P_{ST} \setminus C_{ST}$, and then ‘shave off’ unnecessary moves. To operationalize this, it suffices to define comparative parsimony over the set $\Sigma_{ST}$ of possible strategies for $P_{ST} \setminus C_{ST}$, where $\sigma$ is (comparatively) no less (strictly more) parsimonious than $\sigma'$, noted $\sigma \preceq_p \sigma'$ ($\sigma \succ_p \sigma'$) if $\sigma$ flips at most as many cards as (strictly less cards than) $\sigma'$ in at least one possible $S_0$.

(C5) is an all-or-nothing constraint, asking ‘absolute’ maximization of parsimony, and makes $\preceq_p$-order almost irrelevant save if there is some $\preceq_p$-maximal $\sigma$ that flips no unnecessary cards. But $\preceq_p$ yields a heuristic for finding strategies satisfying (C5), and eliminate dominated ones. Some strategies may be $\preceq_p$-incomparable, but the heuristics only operates on $\preceq_p$-chains, e.g. starts from some $\sigma$ and, varying selections, to induces a sequence $\sigma' \preceq_p \ldots \preceq_p \sigma$ where $\sigma'$ satisfies (C5). We use this heuristics in the rest of this section.

We will first examine how strategies fare w.r.t. $P_{ST} \setminus C_{ST}$, before adding (C5). We begin with the Wason learning function $\sigma_w$, i.e. the one-shot r-c strategy that outputs $\text{turn}[\{A, \cdot\}, \{7, \cdot\}]$, performs a test, and stops. Depending on $S_0$, $\sigma_w$ updates $\text{Inf}^0_w(S_0)$ into one of the following:

\[
\begin{align*}
\text{Inf}^0_w(S_0) &= \{(A, 4), (K, [\frac{3}{4}]), (4, [\frac{3}{4}]), (7, K)\} & \text{Inf}^0_w(S_0) &= \{(A, 7), (K, [\frac{3}{4}]), (4, [\frac{3}{4}]), (7, K)\} \\
\text{Inf}^0_w(S_0) &= \{(A, 4), (K, [\frac{4}{3}]), (4, [\frac{3}{4}]), (7, A)\} & \text{Inf}^0_w(S_0) &= \{(A, 7), (K, [\frac{3}{4}]), (4, [\frac{3}{4}]), (7, A)\}
\end{align*}
\]

Since $\sigma_w$ recommends the same selection in all sixteen possible states $S_0$, $\sigma_w$ is a uniform r-c strategy w.r.t. to selections, even if it is not w.r.t. to its assessment. This sense of uniformity being the only one we need, we can drop the qualification without ambiguity. Notice also that ‘one-shot’-ness implies uniformity, but not the converse.

While $\sigma_w$ decides $P_{ST} \setminus C_{ST}$ simpliciter, it sometimes obtains unnecessary information. However, there is no uniform r-c $\sigma$ deciding $P_{ST}$ s.t. $\sigma \preceq_p \sigma_w$: if $\sigma$ drops either $(A, \cdot)$ or

---

7This mimics elimination of dominated strategy, one of the standard of strategic inference for solving games. Classical game theory assumes that players perform it ‘top-down’ from complete representations including every position the game may reach. Heuristics or intensional tests perform it ‘bottom-up’, and although not critical in ST (whose representation is finite) they are generality constraint for algorithmic models, since complete representations of problems algorithmic problem-solves can address are not always computable [1, 4, 11].

8Uniformity is borrowed from the theory of extensive games of imperfect information, in which players move in turn, and where player $X$ do not always know $Y$’s past moves before choosing her own. A uniform strategy for player $X$ is one that does not include ‘contingency plans’ conditional on $Y$’s past moves. Because learning problems are typically interpreted as extensive games vs. Nature, where ‘Nature’s strategy’ includes the selection of $S_0$ (unknown to Inquirer), uniform strategies are those that do not depend on the state of Nature [7, 10, 12].
(7, -), it cannot apply decisively the test when resp. \((A, 4) \in S_0\) or \((7, K) \in S_0\). Moreover, any uniform \(v\)-c \(\sigma\) that asks more than \(\sigma_w\) select either \((4, .)\) or \((K, .)\), which are unnecessary. Therefore, it also holds that \(\sigma_w \leq_p \sigma\) for any \(\sigma\), and we have:

**Observation 1.** \(\sigma_w\) is the most parsimonious uniform \(v\)-c strategy that decides \(P_{ST} \setminus C_{ST}\).

**Obs. 1** does not generalize to nonuniform \(v\)-c strategies. The nonuniform \(v\)-c \(\sigma_1\) that plays \(\text{turn}[(A, .)]\), tests for \(Q_{ST}\) and halts if successful, otherwise plays \(\text{turn}[(7, .)]\), performs a test, and halts, decides \(P_{ST} \setminus Q_{ST}\) more parsimoniously than \(\sigma_w\) on \(S_0\) when \((A, 7) \in S_0\) and thus \(\sigma_1 \leq_p \sigma_w\). Likewise, \(\sigma_2\), which is just like \(\sigma_1\), but plays \(\text{turn}[(7, .)]\) first, then \(\text{turn}[(A, .)]\) if necessary, beats \(\sigma_w\) on \(S_0\) when \((7, A) \in S_0\), and thus \(\sigma_1 \leq_p \sigma_w\). Moreover, ‘shaving off’ \(\sigma_1\) or \(\sigma_2\) leaves only strategies asking one card or none, which by are uniform and more parsimonious than \(\sigma_w\), and by **Obs. 1** cannot solve \(P_{ST} \setminus C_{ST}\). Therefore, we have:

**Observation 2.** \(\sigma_1\) and \(\sigma_2\) are the most parsimonious \(v\)-c strategies that decides \(P_{ST} \setminus C_{ST}\).

Since neither \(\sigma_1 \leq_p \sigma_2\) nor \(\sigma_1 \leq_p \sigma_2\) hold, none of \(\sigma_1\) or \(\sigma_2\) can takes precedence over the other without e.g. assumption about probabilities of \((A, 7)\) and \((7, A)\)-states. But they both of strictly \(\leq_p\)-dominate \(\sigma_w\), and as consequence **Post. A** does not hold in general in \(P_{ST}\), because \(P_{ST}\) imposes additional constraints that restrict the range of admissible strategies, and cannot ‘reinstate’ \(\sigma_w\). For the same reason, **Obs. 2** does not suffice yet to determine whether \(P_{ST}\) is decidable, but since we already know that \(\sigma_1\) and \(\sigma_2\) are the most parsimonious strategies for \(P_{ST} \setminus C_{ST}\), all we need is to try them against \((C5)\).

\((C5)\) is compatible with both methods to appreciate whether some \((X, \cdot)\) is necessary, both of which assume that the problem is planned before a strategy is actually selected. Hence, the following heuristics correspond to ‘subroutines’ that run simulation, before they actually select a strategy. **Method 1** evaluates the contribution of \((X, \cdot)\) w.r.t. all \(\text{Inf}^0_j(S_0)\) to which the test can performed. **Method 2** is more liberal, and evaluates the contribution of \((X, \cdot)\) w.r.t. the contribution other cards in some \(\text{Inf}^0_j(S_0)\) over which a test is performed. Applied to either \(\sigma_1\) or \(\sigma_2\), **Method 1** evaluates the contribution of resp. \((A, \cdot)\) or \((7, \cdot)\) as unnecessary when \(S_0 \in \overline{\text{Rule}}\) and resp. \((A, 4) \in S_0\) or \((7, K) \in S_0\). As above, using the \(\leq_p\)-based heuristics ‘shaves off’ the first move of both, so that the only admissible strategy is \(\text{turn}[0]\), which leaves \(P_{ST}\) unsolvable under constraints. **Method 2** returns the same verdict, although eliminates both strategies in one fell swoop.\(^9\)

One could insist that \((C5)\) is obviously too strong, and that instructions are (implicitly) ordered lexicographically, and merely impose to select some \(\leq_p\)-preferred \(\sigma\) that still decides \(P_{ST} \setminus C_{ST}\). Let us refer to this ‘weakened’ \((C5)\) as \((C5^*)\), and accordingly let \(P_{ST}^*\) denote the problem identical with \(P_{ST}\) save that \(C_{ST}\) imposes minimization, rather than elimination, of unnecessary cards. With \((C5^*)\), \(\sigma_1\) and \(\sigma_2\) are reinstated as admissible solutions, but \(\leq_p\) remains the main heuristic tool in \(P_{ST}^*\), and since both \(\sigma_1 \leq_p \sigma_w\) and \(\sigma_2 \leq_p \sigma_w\), **Post. A** does not hold for with it either.

---

\(^9\) **Method 2** does not require a complete representation of \(P_{ST}\) (cf. n. 7), and Stenning & Van Lambalgen report experimental evidence that subjects sometimes implement it [18, pp. 59 sq.]. Details are provided below for the technically minded reader, but are unnecessary for the rest of our argument. Assume that evaluation with **Method 1** begins with \(\sigma_1\). As soon as the information structure \(\text{Inf}^0_j(S_0) = \{(A, 4), (K, [\bar{c}]), (4, [\bar{c}]), (7, K)\}\) is computed, \(\text{turn}[(A, .)]\) appears unnecessary in retrospect, because the test is successful even without \((A, .)\). \((A, \cdot)\) must be ‘shaved off’ resulting in \(\text{turn}[0]\), and a switch to \(\text{turn}[(7, .)]\) i.e. a switch to \(\sigma_2\). Then, as soon as an information structure: \(\text{Inf}^0_j(S_0) = \{(A, 7), (K, [\bar{c}]), (4, [\bar{c}]), (7, A)\}\) is computed, a elimination of \((A, .)\) ensues by the above argument (left to the reader), but now, \(\sigma_1\) has already been eliminated and \(P_{ST}\) appears finally unsolvable under constraints. An symmetric argument (left to the reader) beginnings with \(\sigma_2\), leads to the same conclusion.
Given Obs. 1, an additional constraint (C6a), imposing uniform strategies, restores Post. A. One can formally define \( P^w_{ST} \) as being just like \( P_{ST} \) save that \( C_{ST}^w \) adds uniformity to \( C_{ST} \). Uniformity is an all-or-nothing matter: it partitions \( \Sigma_{ST} \), and eliminates \( \sigma_1 \) and \( \sigma_2 \), and since \( \sigma_w \) is acceptable under \( (C5^*) \), and no other constraint applies, Obs. 1 can be strengthened to \( P^w_{ST} \), and Post. A does hold for it. Since uniformity is not explicit in the instructions of ST, and the maneuver does not vindicate Post. A, but treats it as an implicit definition of ‘optimality’ in ST. An arguably better approach cashes in on how \( (C4) \) favors heuristically one-shot strategies (cf. n. 6), rephrasing Post. A as:

**Postulate B.** Under \( (C1)-(C5^*) \), dominant options report exactly \((A, \cdot) \) and \((7, \cdot) \)

Post. B hits closer to the mark, because \( \sigma_1 \) and \( \sigma_2 \) sometimes report both \((A, \cdot) \) and \((7, \cdot) \) as final selections, but does not hold for either \( P_{ST} \) or \( P^w_{ST} \). Interestingly, it holds for \( P^w_{ST} \), and also if one enforces a restriction \( (C6b) \) to one-shot strategies instead of \( (C6a) \), i.e. defining \( P^1_{ST} \) just as \( P^w_{ST} \), save that \( C^1_{ST} \) substitutes a restriction to one-shot strategies instead of uniform ones. Summing up the discussion of this section, we have:

**Claim 3.** ST can be modeled as either \( P_{ST} \), \( P^w_{ST} \), \( P^w_{ST} \) or \( P^1_{ST} \).

1. If ST is \( P_{ST} \), then ST is unsolvable under constraints, and Post. A does not hold;
2. If ST is \( P^w_{ST} \), ST is decidable under constraints, but Post. A does not hold;
3. If ST is \( P^w_{ST} \) or \( P^1_{ST} \), ST is decidable under constraints, and Post. A holds. \( \square \)

4 Discussion: How deductive is ST?

\( P_{ST} \setminus C_{ST} \) is decidable by \( r-c \) strategies, which also suffice to decide mechanically propositional and syllogistic problems. These problems are solvable with finite (mechanizable) proofs, and any \( r-c \) strategy waiting for the completion of a mechanized proof (and for its final verdict) decides any such problem. However, being decidable by \( r-c \) strategies is neither a necessary nor a sufficient condition of for being a deductive problem. It is not necessary, because propositional logic and syllogistics are decidable fragments of first-order logic, which is not decidable by \( r-c \) strategies (proofs are mechanizable but not always finite).

It is not sufficient either, because uniformity and parsimony are preferential criteria, and optimizing their satisfaction makes ST (and any of its decidable variants) decision-making problems, which are not typically though of as ‘purely’ deductive.

One can propose to revise this last identifications, and reduce logicality to the possibility of logical treatment. Indeed, although \( P_{ST} \) is unsolvable, its solvability is a mechanically decidable problem, because it requires only a finite procedure (that we have partially implemented in this paper). Solvability of any of \( P^w_{ST} \), \( P^w_{ST} \) and \( P^1_{ST} \) is also decidable mechanically, because it requires no more tests than the solvability problem for \( P_{ST} \). All these problems can be addressed by first-order theorem-provers (or model-checkers) in finite time, and amenability to such logical treatment is sufficient to call ST, \( ST^* \), \( ST^w \) or \( ST^1 \) ‘logical.’ But we cannot stop there unless we beg the question of logicality: we could as well call them ‘algebraic’ or ‘electrical’ (computer code ultimately reduces to Boolean algebra, realized by electrical circuits). We should resist this conclusion, which confuses the phenomenon modeled with the formal (and mechanical) tools used to model it.

Theorists who downplay the role of classical deduction in reasoning, still accept the content of Post. A (or Post. B), have no such sophisticated reasons. Some Bayesians still profess a blind faith in the normative status of logic— “[l]ogically, subjects should select

---

10 A mechanical solution is possible, but is not finite, requires an impatient method (structurally equivalent to the one that solves the halting problem n.) and is inductive, rather than deductive [11].
only the P and not-Q cards” [13, p. 608] (our emphasis)—only mitigated by the contention ST is “a ‘loose’ probabilistic task, rather than a ‘tight’ deductive task” [13, p. 626]. Relevance theorists offer a better than this ‘loose’ talk, i.e. insights on causes for the logical (mis)conception of ST as a ‘deductive’ task:

Since the rule is true […] unless there are items combining the P and the not-Q properties, the logically correct selection is that of the P and the not-Q cards, each of which could turn out to provide a counter-example. [5, B71]

(Emphasis added.)

This quotes shows a drift from the truth-functional interpretation of if…then…, to valid propositional inference schemas that conditional warrant, possibly under the impression that the latter should be applied no matter what, because logic is ‘abstract.’ But even so conceived, logic is a medium of computation, and what is to be computed is a decision problem. Relying on heuristics based on the meaning of the conditional is useless ST, does not solve ST∗, and only solve ST1 and ST2 through ‘inferential luck,’ ignoring the relevant constraints (parsimony and uniformity). Arguing that these inference patterns coordinates subjects with experimenters amounts to asking subjects to coordinate on the same misconception as theorists. Empirical implementation of ST shows that human reasoners are at least proficient enough not to do so systematically.

5 Conclusion

Bayesians and Relevance Theorists deserve credit for having renewed the study of ST, and this paper contribute to the line of investigation they opened. The Bayesian theory boils down to the hypothesis that, in abstract and thematic, variants of ST, subjects systematically fall back to sampling and a ‘cheater-detection’ (and perform optimally). In the latter performance matches expectations, but amounts to adopt uniform strategies, possibly as result of ‘worst case’ reasoning in social contexts.11 Pace Bayesians, other factors than social values can favor parsimonious uniform v-c strategies. Quite ironically, the feature of our model supporting this conclusion generalizes the Bayesian account of ‘deontic’ cases: Bayesian add to the conditional a utility function, and utilities are cardinal expressions of preference orderings (or the aggregation of several such orderings), that we have shown relevant to ‘abstract’ cases.

Our model can already accommodate Bayesian ‘rational analysis’ of the ‘sampling’ interpretation, and nothing in principle prevents it to accommodate other interpretations favoring P-and-Q selections. An interesting question is whether it can account for the data produced by Relevance theorists. Contrary to Bayesians, who merely analyzed old data, Relevance theorists have produced manipulation in which “the same rule, regardless of whether it is tested descriptively or deontically, can be made to yield more P-and-Q selections or more P-and-not-Q selections” [5, B70].

Relevance Theorists take subjects’ lack of awareness of manipulations as evidence that pragmatic effects trump semantic treatment, yet qualify pragmatic factors in very general terms: assessment of relevance to the task at hand, and cost of inference to reveal relevance. Our model allows for a finer-grained characterization of relevance as semantic and communicational salience, for which awareness is not necessary, when semantic treatment can rest on ‘precomputations’ inherited from past linguistic interaction [18, pp. 112-113]. Coordination on relevant-qua-salient features of instructions is a precondition for subjects’ performance to match expectations of theorists.

11Such reasoning selects strategies either equivalent to σ∗u, or strategies equivalent to σ∗1 and σ∗2 which are just like σ1 and σ2, save that they always play the second second move. Although σi ≺p σj for all i, j ∈ {1, 2}, uniformity overrules parsimony, because in the worst case all those who can violate the rule actually do so.
Lack of coordination due to flawed design is not unique to ST, and J. Jacot has recently argued that it is predictable [8]. She analyses the double disjunction task (from [9]) where e.g. subjects are presented with premises like: ‘Alice is in India, or Charles is in Afghanistan, but not both’ and: ‘Barbara is in Pakistan, and Charles is in Afghanistan, but not both,’ then asked to draw propositional conclusions. Jacot argues that this amounts to identify as salient syntactic features (connectives), but that subjects are more likely to select as other semantic features (alternative locations of individuals), in an attempt to convey useful information.\footnote{The task has been replicated by Bouquet & Warglien [2] based on a finer model of information processing, which identifies salience features more accurately. Subsequently, performance matches predictions. Although not analyzed in the same pragmatic terms, these results support Jacot's analysis.}

Ironically, the disrupting storyline was inserted to offset consequences of ‘abstract’ instructions, under the misconception (inherited from ST) that thematic tasks are ‘deductively’ easier. Results in the original disjunctive tasks have been used to formulate far-reaching conclusions about human reasoning abilities [15, 19], that stand and fall with their ‘logical’ underpinnings. Unsophisticated conceptions of logic, semantics, and pragmatics, has not merely ‘rigged the ST game’ against coordination, but have for too long inspired sub-par experimental designs and ill-grounded interpretations of data. We hope to have contributed to counteract that influence.

Acknowledgments. Research for this paper was supported by a postdoctoral grant from the Swedish Research Council, and is part of an ongoing joint work with Justine Jacot, Ph.D. candidate in philosophy (Lund). A former co-authored version was presented at the Logic & Cognition Conference, in Poznan (May 2012). We thank participants for their comments. I am responsible for all additions and modifications, and accept sole responsibility for errors I could have introduced. Justine Jacot deserves at least as much credit as myself for everything else.

References


