Information Processing and Constraint Satisfaction in Wason's Selection Task

Genot, Emmanuel

Published in:
Cognition, reasoning, emotion, Action. CogSc-12. Proceedings of the ILCLI International Workshop on Cognitive Science.

2012

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** $\boxed{\text{A}}\boxed{\text{K}}\boxed{4}\boxed{7}$

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*,[1] 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 $(A, \cdot)$ and the *Q*-card $(7, \cdot)$ (where in (X, Y), X is the initially visible face first, and '$\cdot$' an unknown value). In the above variant, however, less than 10% of subjects conform to expectations, with as typical results: $(A, \cdot)$ alone 35%; $(A, \cdot)$ and $(4, \cdot)$, 45%; $(A, \cdot)$ and $(7, \cdot)$, 5%; $(A, \cdot)$, $(4, \cdot)$ and $(7, \cdot)$, 7%; and other selections, 8% [18]. [2]

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 $(A, \cdot)$ and $(K, \cdot)$ (resp.: $(4, \cdot)$ and $(7, \cdot)$) **(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,

[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* $(A, \cdot)$ *and* $(7, \dot{})$ *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* (IMI) propose by J. Hintikka and his associates [6, 7]. The IMI 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 $\mathrm{SoN}(T)$. The IMI is a special case of (algorithmic) learning models, in which solvability of $\langle T, Q \rangle$ 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 IMI, σ additionally affects how the data sequence is collected. Whenever some state $S \in \mathrm{SoN}(T)$ satisfies some $q_i \in Q$ (in symbol: $S \models q_i$), σ *solves* $\langle T, Q \rangle$ *in* S iff for every sequence of observations characterizing S, there is a finite segment after which σ outputs $q_i \in Q$ and never changes its assessment. Accordingly, σ *solves* $\langle T, Q \rangle$ *simpliciter* iff σ solves $\langle T, Q \rangle$ in every $S \in \mathrm{SoN}(T)$.

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 $q_i \in Q$ 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 $\mathrm{SoN}(T)$, 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 $\mathrm{SoN}(T)$, 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 $q_i \in Q$, 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 $q \in Q$ is finite, and all the values for those parameters are available in S, $\langle T, Q \rangle$ 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].[4] For a decidable problem, a *patient* σ that exhausts all and only the relevant parameters always solves (and decides, if halting) a problem no later than an impatient σ' identical with σ save possibly for the initial assessment, an sometimes earlier if σ' must revise it, i.e. *ceteris paribus* a patient σ (weakly) *dominates* an impatient σ' 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 σ can still *halt later than success*. In particular, if in some P the overall set of parameters for S is finite, if σ exhaustively enumerates them, σ *solves* P no later (*ceteris paribus*) than an impatient or cautious σ', but may *decide* P later than σ' if σ' collects only sufficiently many parameters to solve P and halts on success, while σ is still collecting. Some constraints on strategies can sometimes optimize solution (or decision), and a *constrained learning problem* P is a triple $P = \langle T, Q, C\rangle$, where C is a set of constraints. Notice that, when the constraints are too strong, P may not be solvable even it P minus C (noted $P \setminus C$) is .

3 Results

The intended interpretation of ST given by **(C1)–(C5)** characterizes a *constrained learning problem* $P_{\mathrm{ST}} = \langle T_{\mathrm{ST}}, Q_{\mathrm{ST}}, C_{\mathrm{ST}}\rangle$, whose properties are independent on whether the problem solver is, human, mechanical or abstract, has awareness or not, etc.[5] **(C1)** is equivalent to the part of T_{ST} that specifies (four) relevant parameters of every $S \in \mathrm{SoN}(T_{\mathrm{ST}})$ as follows:

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

We let $S_0 \in \mathrm{SoN}(T_{\mathrm{ST}})$ 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:

$$\forall\sigma, \quad \mathsf{Inf}_0^{\sigma}(S_0) = \{(A, \left|{}^{4}_{7}\right|), (K, \left|{}^{4}_{7}\right|), (4, \left|{}^{A}_{K}\right|), (7, \left|{}^{A}_{K}\right|)\}\sigma$$

where in $(X, \left|{}^{y_1}_{y_2}\right|)$, X is known, and there is uncertainty between y_1 and y_2. $\mathsf{Inf}_0^{\sigma}(\mathrm{SoN})$ can be expanded to a truth-table-like 16-rows matrix representing the *equivalence class* of states indiscernible from S_0, identical with $\mathrm{SoN}(T)$ 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 $\mathtt{turn}[X]$ 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.[6] **(C2)** also implies that the information about the back of cards is reliable (at least as much as a visual check is).**(C3)** bipartitions $\mathrm{SoN}(T)$, through the truth-functional meaning of *if... then* as follows:

$$\mathsf{Rule} = \{\{(A, x_1), (K, x_2), (4, x_3), (7, x_4)\} : x_1 = 4 \text{ and } x_4 = K\} \tag{Rule}$$

$$\overline{\mathsf{Rule}} = \{\{(A, x_1), (K, x_2), (4, x_3), (7, x_4)\} : x_1 = 7 \text{ or } x_4 = A\} \tag{$\overline{\mathsf{Rule}}$}$$

[4]A paradigmatic undecidable problem is the *halting problem* for a program that runs either finitely or infinitely many steps, and asks for a σ 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' σ 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.

[5]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)

[6]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 $Q_{\text{ST}} = \{\mathsf{Rule}, \overline{\mathsf{Rule}}\}$, because characteristic properties of Rule and $\overline{\mathsf{Rule}}$ 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 $\overline{\mathsf{Rule}}$ and output an answer if possible, otherwise return '?'", because **(C1)**–**(C3)** and the reliability of (outputs of) selections entail that *patient-credulous strategies (weakly) dominate* (*ceteris paribus*) in P_{ST}. No strategy can then satisfy **Post. A** *and* strictly dominate a P-C σ that also satisfies it, and therefore we restrict our attention to only those.

The first explicit constraint is **(C4)** imposing *decision*. Since $P_{\text{ST}} \setminus C_{\text{ST}}$ 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, $\mathsf{Inf}_0^{\sigma}(S_0)$ must shrink down to at least:

$$\mathsf{Inf}_i^{\sigma}(S_0) = \begin{cases} \{(A,4),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,K)\} & \text{if } S_0 \in \mathsf{Rule};\\ \{(A,7),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|)\} \text{ or } \{(A,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,A)\} & \text{if } S_0 \in \overline{\mathsf{Rule}} \end{cases}$$

(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 σ, check whether it solves $P_{\text{ST}} \setminus C_{\text{ST}}$, and then 'shave off' unnecessary moves.[7] To operationalize this, it suffices to define *comparative parsimony* over the set Σ_{ST} of possible strategies for $P_{\text{ST}} \setminus C_{\text{ST}}$, where σ is (comparatively) *no less* (*strictly more*) *parsimonious than* σ', noted $\sigma \precsim_p \sigma'$ ($\sigma \prec_p \sigma'$) if σ flips at most as many cards as (strictly less cards than) σ' in at least one possible S_0.

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

We will first examine how strategies fare w.r.t. $P_{\text{ST}} \setminus C_{\text{ST}}$, before adding **(C5)**. We begin with the *Wason learning function* σ_w, i.e. the one-shot P-C strategy that outputs $\mathtt{turn}[(A,\cdot),(7,\cdot)]$, performs a test, and stops. Depending on S_0, σ_w updates $\mathsf{Inf}_0^{\sigma_w}(S_0)$ into one of the following:

$$\mathsf{Inf}_1^{\sigma_w}(S_0) = \{(A,4),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,K)\} \qquad \mathsf{Inf}_2^{\sigma_w}(S_0) = \{(A,7),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,K)\}$$
$$\mathsf{Inf}_3^{\sigma_w}(S_0) = \{(A,4),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,A)\} \qquad \mathsf{Inf}_4^{\sigma_w}(S_0) = \{(A,7),(K,\left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|),(4,\left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|),(7,A)\}$$

Since σ_w recommends the same selection in all sixteen possible states S_0, σ_w is a *uniform* P-C *strategy* w.r.t. to selections, even if it is not w.r.t. to its assessment.[8] 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 σ_w decides $P_{\text{ST}} \setminus C_{\text{ST}}$ *simpliciter*, it sometimes obtains unnecessary information. However, there is no uniform P-C σ deciding P_{ST} s.t. $\sigma \precsim_p \sigma_w$: if σ drops either $(A,\cdot)$ or

[7] This 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].

[8] *Uniformity* 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].

$(7, \cdot)$, it cannot apply decisively the test when resp. $(A, 4) \in S_0$ or $(7, K) \in S_0$. Moreover, any uniform P-C σ that asks more than σ_w select either $(4, \dot{\ })$ or $(K, \dot{\ })$, which are unnecessary. Therefore, it also holds that $\sigma_w \precsim_p \sigma$ for any σ, and we have:

Observation 1. *σ_w is the most parsimonious uniform* P-C *strategy that decides $P_{ST} \setminus C_{ST}$.*

Obs. 1 does not generalize to *nonuniform* P-C strategies. The nonuniform P-C σ_1 that plays $\mathtt{turn}[(A, \cdot)]$, tests for Q_{ST} and halts if successful, otherwise plays $\mathtt{turn}[(7, \cdot)]$, performs a test, and halts, decides $P_{\mathrm{ST}} \setminus Q_{\mathrm{ST}}$ more parsimoniously than σ_w on S_0 when $(A, 7) \in S_0$, and thus $\sigma_1 \prec_p \sigma_w$. Likewise, σ_2, which is just like σ_1, but plays $\mathtt{turn}[(7, \cdot)]$ first, then $\mathtt{turn}[(A, \cdot)]$ if necessary, beats σ_w on S_0 when $(7, A) \in S_0$, and thus $\sigma_1 \prec_p \sigma_w$. Moreover, 'shaving off' σ_1 or σ_2 leaves only strategies asking one card or none, which by are uniform and more parsimonious than σ_w, and by **Obs. 1** cannot solve $P_{\mathrm{ST}} \setminus C_{\mathrm{ST}}$. Therefore, we have:

Observation 2. *σ_1 and σ_2 are the most parsimonious* P-C *strategies that decides $P_{ST} \setminus C_{ST}$.*

Since neither $\sigma_1 \precsim_p \sigma_2$ nor $\sigma_1 \precsim_p \sigma_2$ hold, none of σ_1 or σ_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 $\precsim_p$-dominate σ_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' σ_w. For the same reason, **Obs. 2** does not suffice yet to determine whether P_{ST} is decidable, but since we already know that σ_1 and σ_2 are the most parsimonious strategies for $P_{\mathrm{ST}} \setminus C_{\mathrm{ST}}$, all we need is to try them against **(C5)**.

(C5) is compatible with two 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* $\mathsf{Inf}_j^{\sigma}(S_0)$ 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 $\mathsf{Inf}_i^{\sigma}(S_0)$ over which a test is performed. Applied to either σ_1 or σ_2, *Method 1* evaluates the contribution of resp. $(A, \cdot)$ or $(7, \cdot)$ as unnecessary when $S_0 \in \overline{\mathsf{Rule}}$ and resp. $(A, 4) \in S_0$ or $(7, K) \in S_0$. As above, using the $\precsim_p$-based heuristics 'shaves off' the first move of both, so that the only admissible strategy is $\mathtt{turn}[\emptyset]$, 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 $\precsim_p$-preferred σ that *still decides* $P_{\mathrm{ST}} \setminus C_{\mathrm{ST}}$. 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*)**, σ_1 and σ_2 are reinstated as admissible solutions, but $\precsim_p$ remains the main heuristic tool in P^*_{ST}, and since both $\sigma_1 \prec_p \sigma_w$ and $\sigma_2 \prec_p \sigma_w$, **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 σ_1. As soon as the information structure $\mathsf{Inf}_i^{\sigma_1}(S_0) = \{(A, 4), (K, \left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|), (4, \left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|), (7, K)\}$ is computed, $\mathtt{turn}[(A, \cdot)]$ appears unnecessary in retrospect, because the test is successful even without $(A, \cdot)$. $(A, \cdot)$ must be 'shaved off' resulting in $\mathtt{turn}[\emptyset]$, and a switch to $\mathtt{turn}[(7, \cdot)]$ i.e. a switch to σ_2. Then, as soon as an information structure: $\mathsf{Inf}_j^{\sigma_2}(S_0) = \{(A, 7), (K, \left|\begin{smallmatrix}4\\7\end{smallmatrix}\right|), (4, \left|\begin{smallmatrix}A\\K\end{smallmatrix}\right|), (7, A)\}$ is computed, a elimination of $(A, \cdot)$ ensues by the above argument (left to the reader), but now, σ_1 has already been eliminated and P_{ST} appears finally *unsolvable under constraints.* An symmetric argument (left to the reader) beginnings with σ_2, leads to the same conclusion.

Given **Obs. 1**, an additional constraint **(C6a)**, imposing *uniform* strategies, restores **Post. A**. One can formally define P^u_{ST} as being just like P^*_{ST} save that C^u_{ST} adds uniformity to C^*_{ST}. Uniformity is an all-or-nothing matter: it partitions Σ_{ST}, and eliminates σ_1 and σ_2, and since σ_w is acceptable under **(C5*)**, and no other constraint applies, **Obs. 1** can be strengthened to P^u_{ST}, and **Post. A** does hold for it. Since niformity 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* $(7,\dot{)}$

Post. B hits closer to the mark, because σ_1 and σ_2 sometimes report both $(A,\cdot)$ and $(7,\dot{)}$ as final selections, but does not hold for either P_{ST} or P^*_{ST}. Interestingly, it holds for P^u_{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^u_{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^*_{ST}, P^u_{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^*_{ST}, ST is* decidable *under constraints, but* ***Post. A*** *does not hold;*
3. *If ST is P^u_{ST} or P^1_{ST}, ST is* decidable *under constraints, and* ***Post. A*** *holds.* □

4 Discussion: How deductive is ST?

$P_{\mathrm{ST}} \setminus C_{\mathrm{ST}}$ is decidable by P-C strategies, which also suffice to decide mechanically propositional and syllogistic problems. These problems are solvable with *finite* (mechanizable) proofs, and any P-C strategy waiting for the completion of a mechanized proof (and for its final verdict) decides any such problem. However, being decidable by P-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 P-C strategies (proofs are mechanizable but not always finite).[10] 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^*_{ST}, P^u_{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^u 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 ST^u and ST^1 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 P-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.

[11] Such reasoning selects strategies either equivalent to σ_w, or strategies equivalent to σ_1^* and σ_2^* which are just like σ_1 and σ_2, save that they *always* play the second second move. Although $\sigma_i \prec_p \sigma_j^*$ for all $i, j \in \{1, 2\}$, uniformity overrules parsimony, because in the worst case all those who *can* violate the rule *actually* do so.

Lack of coordination due to flawed disign 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 presents 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.[12]

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

[1] 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.

[2] Paolo Bouquet and Massimo Warglien. Mental models and Local models semantics: the problem of information integration. In *Proceedings of the European Conference on Cognitive Science (ECCS'99)*, pages 169–178. University of Siena, Italy, 1999.

[3] Emmanuel J. Genot. The Game of Inquiry. The Interrogative Approach to Inquiry and Belief Revision Theory. *Synthese*, 171:271–289, 2009.

[4] Emmanuel J. Genot and Justine Jacot. Semantic Games for Algorithmic Players. To appear. Presented at the 6th Workshop on Decision, Games and Logic, June 2012, Munich (Germany), 2012.

[5] 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.

[6] Jaakko Hintikka. *Socratic Epistemology*. Cambridge University Press, Cambridge, 2007.

[12] 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.

[7] Jaakko Hintikka, Ilpo Halonen, and Arto Mutanen. Interrogative Logic as a General Theory of Reasoning. In *Inquiry as Inquiry*, pages 47–90. Kluwer, Dordrecht, 1999.

[8] Justine Jacot. The Double Disjunction Task as a Coordination Problem. In Jakub Szymanik and Rineke Verbrugge, editors, *Proceedings of the Logic & Cognition Workshop at ESSLLI 2012, Opole, Poland, 13-17 August, 2012*, volume 883 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.

[9] P.N. Johnson-Laird, Ruth M.J. Byrne, and Walter Schaeken. Propositional Reasoning by Model. *Psychological Review*, 99(3):418–439, 1992.

[10] Kevin T. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, New York, 1996.

[11] Kevin T. Kelly. Uncomputability: the problem of induction internalized. *Theoretical Computer Science*, 317(1–3):227–249, 2004.

[12] Eric Martin and Daniel Osherson. Scientific Discovery Based on Belief Revision. *Journal of Symbolic Logic*, 62:1352–1370, 1997.

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

[14] Karl R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959.

[15] Eldar Shafir. Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, 50:403–430, 1994.

[16] Keith Stenning and Michiel van Lambalgen. Semantics as a Foundation for Psychology: A Case Study of Wason's Selection Task. *Journal of Logic, Language and Information*, 10:273–317, 2001.

[17] Keith Stenning and Michiel van Lambalgen. The Natural History of Hypotheses About the Selection Task Reasoning. In Ken Manktelow and Man Cheung Chung, editors, *Psychology of Reasoning: Historical and Theoretical Perspectives*, chapter 6, pages 126–156. Psychology Press, Hove, New York, 2004.

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

[19] Maggie E. Toplak and Keith E. Stanovich. The Domain Specificity and Generality of Disjunctive Reasoning: Searching for a Generalizable Critical Thinking Skill. *Journal of Educational Psychology*, 94:197–209, 2002.

[20] Peter Cathcart Wason. Reasoning about a rule. *Quaterly Journal of Experimental Psychology*, 20(3):273–281, 1968.