# LUND UNIVERSITY

## Taking Problem-Solving Seriously

Genot, Emmanuel; Jacot, Justine

2012

# Taking Problem-Solving Seriously:
# A Learning-Theoretic Analysis of the Wason Selection Task

**Emmanuel J. Genot**, **Justine Jacot**[*]

*University of Lund, Department of Philosophy, Kungshuset, Lundagård, 222 22 Lund (Sweden)*

**Abstract.** Instructions in Wason's *Selection Task* underdetermine empirical subjects' representation of the underlying problem, and its admissible solutions. We model the Selection Task as an (ambiguous) *interrogative learning problem*, and reasoning to solutions as: (*a*) selection of a representation of the problem; and: (*b*) strategic planning from that representation. We argue that recovering Wason's 'normative' selection is possible only if both stages are constrained further than they are by Wason's formulation. We conclude comparing our model with other explanatory models, w.r.t. to empirical adequacy, and modeling of bounded rationality.

## 1  Introduction

For thirty years, the four-card *Selection Task*—hereafter "the ST" (see e.g. Wason 1968)—was interpreted as showing that systematic reasoning biases get in the way of *deductive* reasoning. In the 1990s, Bayesian models of rational analysis reinterpreted the data as evidence for (successful) *inductive* reasoning in the ST. Both analyses took deductive reasoning to coincide with *classical* logic, a contention undermined by contemporary developments of non-classical logics (and their formal semantics) largely ignored in the problem-solving literature.

Stenning and van Lambalgen (2008) have proposed a more sophisticated analysis of the ST, based on semantic characterizations of nonclassical logics. It reveals ambiguities in the instructions, underdetermining the representation of the underlying problem (and of its admissible solutions). This analysis explains the variance of subjects' selections by how the (perceived) semantic content of instructions interplays with contextual factors, to induce particular inference patterns. It explains cases problematic for the standard (Bayesian) rational analysis model. Yet, it falls short from generalizing rational analysis, for want of a formal model of reasoning 'to' and 'from' an interpretation, that would explain the heuristic value of semantic reasoning.

---

[*][emmanuel.genot/justine.jacot]@fil.lu.se

This paper outlines such a model, based on a learning-theoretic generalization of Hintikka's game-theoretic *interrogative model of inquiry* (ɪᴍɪ), which shares with other learning-theoretic framework a representation of inquiry as a 'game against Nature' (see Hintikka et al. 1999, Kelly 2007, Martin and Osherson 1998). We argue that strategic planning in interrogative games can capture reasoning in the ST, when *logical reasoning supervenes on, and is continuous with, semantic reasoning*—in particular linguistic interpretation of the instructions. Section 2 provides the background on the ST (2.1), and our formal framework (2.2). Section 3 applies the model to represent the ST (3.1) and examines its possible solutions (3.2). We conclude with a discussion of some empirical results (4.1) and conceptual issues (sec. 4.2).

## 2 Background

### 2.1 "The Mother of All Reasoning Tasks"

Stenning and van Lambalgen (hereafter S&L) call Wason's *Selection Task* (ST) the 'Mother of All Reasoning Task', and give the following formulation for its 'abstract' (original) version:

> 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. (S&L 2008, p. 44)

> **Rule** If there is a vowel on one side, then there is an even number on the other side.
> **Cards** $\boxed{A}$ $\boxed{K}$ $\boxed{4}$ $\boxed{7}$

Neglecting quantification over letters, numbers and sides, **Rule** simplifies in a material conditional: *If P, then Q*. Based on truth-conditions for material conditional, Wason assumed that the 'normative' selection is:

$$(A, \cdot), (7, \cdot) \tag{Nor}$$

where each pair denotes a card, face initially visible first ('·' stands for an unknown value). However, less than 10% of empirical subjects typically conform to (Nor).[1] (Nor)-like selection is higher in 'thematic' variants—where the rule is given a less abstract content. Typical explanations assume that (Nor) is 'the' *context-independent logically competent*—or simply 'logical'—answer, but that: (*a*) in *abstract* tasks, context-sensitive mechanisms override logical competence; and: (*b*) in *thematic* variants, either familiarity helps recover it, or context-sensitive mechanisms emulates it for other reasons.

An influential example is Oaksford and Chater's (hereafter OC) Bayesian *rational analysis* (see O&C 1994). Rational analysis derives the optimal behavior of a cognitive system, under environmental constraints, from specific hypotheses about its goal(s) and computational limitations. A cognitive system can then be contextually optimal even if its output does not obey context-independent

---

[1]Typical results for the abstract version, as reported by S&L 2008, are: $(A, \cdot)$, 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%.

standards. OC assumed that, in the *abstract* task, subjects treat cards as a sample, and **Rule** as an inductive hypothesis, and showed that selections coincided with statistically *optimal data selection*.[2] For the *thematic* task, OC assumed that conditional were read deontically, i.e. as rules that can be (dis)obeyed. That subjects tend to review exhaustingly potential violators (resulting in (Nor)-like selections) reveals strong preferences for the rule not to be violated.

However, as pointed out by S&L 2008, developments in contemporary logical theory (and especially formal semantics) undermine the hypothesis of a single context-independent logical competence. Moreover, empirical evidence casts some doubt on the assumption that, within a category of ST (abstract/thematic), subjects are always doing the same thing (see tutorial dialogs reported in S&L 2008.) On the contrary, selections vary with the *perceived goal*, which can be manipulated by facilitating or hindering particular interpretations of the instructions.

## 2.2    Interrogative learning: a formal model

Let us formally define a *problem* as a triple $P = \langle \mathbf{K}, H, Q \rangle$, where $\mathbf{K}$ is a set of **states** s.t. any $\kappa \in \mathbf{K}$ is (uniquely) identified by values for a set of relevant parameters; $H$ is a set **histories**—sequences of positions at which some value for some parameters are obtained or estimated (gradually identifying an underlying state); and $Q$ is the **principal question** of the problem, which is typically a partition of states in $\mathbf{K}$, although sometimes overlapping answers can be considered.

$H$ corresponds to the view of an omniscient modeler: for every $h \in H$, there is some learning method (see below) addressing $P$ that generates it. Constrains on $H$ apply to *all* LM *for P*. We say that $P$ is *finite* iff $H$ is finite; and with *finite horizon* iff every $h \in H$ is of finite length. If $h$ is of length $n$, $h'$ *extends* $h$ iff $h'|n = h$, where $h'|n$ is the initial segment of length $n$ of $h'$; $h(n)$ denotes the item occurring at the $n$th position of $h$; and $h \frown e$ is the extension of $h$ with $e$.

An (interrogative) *learning method* (LM) for some $P = \langle \mathbf{K}, H, Q \rangle$ is a (partial) function $l : H \longmapsto (\mathbf{A} \cup \mathbf{Q}) \times Q \cup \{?\}$, where $\mathbf{A}$ is a set of (noninterrogative) *actions*; $\mathbf{Q}$ is a set of *instrumental questions*; and '?' indicates suspension between answers in $Q$. $l(H)$ is the subset of $H$ generated implementing $l$ in $P$. A family $L = \{l_1, \ldots, l_n\}$ can induce a *partial representation* $l_1(H) \cup \cdots \cup l_n(H) \subseteq H$ of $P$. We say that $h \in H$ is: (*a*) *l*-terminal iff $l$ is undefined for extensions of $h$; and: (*b*) *maximal* iff it is *l*-terminal for *every l*. When source(s) of answer is (are) 'strategic,' the model covers strategic reasoning about games, since answering strategies can be anticipated by considering constraints on histories.[3] For some problem $P = \langle \mathbf{K}, E, Q \rangle$, interrogative LM $l$, and *l*-terminal $h \in H$:

---

[2]The standard ratios of vowels and even numbers over letters and numbers induces the following ordering: $(A, \cdot) > (4, \cdot) > (7, \cdot) > (K, \cdot)$ (see O&C 1994, p. 625).

[3]The model generalizes Martin and Osherson's *First-Order Paradigm*, and Hintikka's *Interrogative-deductive games*. Both represent problems in some first-order language $\mathcal{L}$, and take $\mathbf{K}$ to be the set of models of some set $\Gamma \subseteq \mathcal{L}$ of premises. Every $\kappa \in \mathbf{K}$ is identified by a *complete diagram* (set of literals, i.e. atomic sentences and their negations). In the first model, parameters are always literals subformulas of elements of $\Gamma$, and for any $l$, $\mathbf{A} = \emptyset$ and $\mathbf{Q}$ includes all *yes-or-no* questions about literals (from the vocabulary of $\Gamma$); additionally, any $l$ can generate a complete diagram, i.e. Nature *eventually answers all atomic questions*. In Hintikka's model, parameters are arbitrary subformulas of elements of $\Gamma$; and for any $l$, $\mathbf{A}$ includes 'semantic actions' to

- *l stabilizes on q in h* at $h(n)$ iff $l(h|n) = \langle X, q \rangle$ for some (possibly empty) $X \subseteq \mathbf{A} \cup \mathbf{Q}$, and: (*a*) *h* is of length $m \geq n$, and for any $n'$ s.t. $n \leq n' \leq m$, $l(h|n') = \langle X', q \rangle$, or: (*b*) *h* is infinite, and for any $n' > n$, $l(h|n') = \langle X', q \rangle$ (for some possibly empty $X' \subseteq \mathbf{A} \cup \mathbf{Q}$).
- *l solves P on h* at $h(n)$ iff (*i*) *l* stabilizes on *q* at $h(n)$ and: (*b*) any $\kappa \in \mathbf{K}$ compatible with values of parameters obtained in *h* is in *q*;
- *l decides P on h* iff *h* is of *finite* length *n*, and *l solves P on h* at $h(m)$ for some $m \leq n$.
- *l* solves (decides) *P* on **K**—or solves *P simpliciter*—iff: (*a*) *l* solves (decides) *P* on every *l*-terminal $h \in H$; and: (*b*) every $\kappa \in \mathbf{K}$ is (partially) characterized in at least one *l*-terminal $h \in l(H)$

Informally, *l* stabilizes in one history *h* whenever it outputs the same answer from some position on, until it stops generating new positions (if it does); solves *P* (in *h*) iff it stabilizes, and gets the answer right (in *h*); and decides *P* (in *h*) whenever it solves it and stops doing anything (immediately, or later). Finally, *l* solves (or decides) *P simpliciter* if *l* always gets the answer right, and examines enough histories to get it right in all states (even under uncertainty about the actual state).

By the above definitions, *halting on success* is only a special case of decision: *l* may stabilize and still output actions (e.g. 'control' questions), before it stops. Therefore, a LM for *P* can solve *P* without deciding it—"no bell rings" when *P* is solved, i.e. when *l* stabilizes (see Kelly 2007). Another consequence is that *any solvable problem with finite horizon is decidable*: additional 'nonredundancy' constraints on LM (e.g. halting on success, etc.) are of special interest for such problems.

Instructions in the ST have several possible interpretations (see below). Such cases can be captured through *generalized problems*, associating to a set *I* of instructions a family $\mathbf{P}_I = P_1, \ldots, P_n$ of possible representations (corresponding to possible interpretations), including a designated problem $P_k^a$, the intended interpretation of *I*, or *actual* problem. A 'dynamic' representation of $P_k^a$ for an agent *X* can be obtained with an *awareness function*, mapping every $h \in H_{P_k^a}$ to a set $\mathbf{P}_I'$ of problems compatible with *X*'s current representation of $P_k^a$.[4] *X*'s representation can be (strictly) partial, if $\mathbf{P}' = \{P_j\}$, and $H_{P_j} \subset H_{P_k^a}$; ambiguous, if $\mathbf{P}'$ is not a singleton; or incorrect, if none of her representations in $\mathbf{P}'$ is a partial representation of $P_k^a$. In problem-solving experiments reasoning to the *intended* interpretation $P_k^a$ of *I* is a *coordination problem* between subject and experimenter.

## 3    The Selection Task as a Learning Problem

### 3.1    Interpreting the instructions

Reasoning *to* an interpretation of the instructions of § 2.1 can be done using them heuristically to select a family of LM *L*, that in turn yields a (partial) representation of the problem. If the representa-

---

analyze elements of Γ in subformulas; and **Q**, both *yes-or-no* questions, and questions with presuppositions obtained by actions in **A**. Also, not all answers are available (even in the limit).

[4]Formally, $P_i' = \langle \mathbf{K}', H', Q' \rangle \in \mathbf{P}'$ iff there is a $P'' = \langle \mathbf{K}'', H'', Q'' \rangle$ s.t. every history in $H''$ is the initial segment of some history in $H'$, and $H'' = l(H_i)$ for some LM *l* that *X* is considering heuristically. Awareness functions were introduced in game theory by Halpern and Rego (2006), to model players reasoning from partial or incorrect representations of games they play.

tion is ambiguous, one then has to find a 'best' candidate. Reasoning *from* an interpretation requires yet another heuristic use of the instructions, inducing a preference ordering over LM that solve the problem (if any). An essential constraint is:

<div align="center">

*Don't turn unnecessary cards*  (NEC)

</div>

Lastly, subjects must report the selection according to the 'best' LM. One should expect subjects to either go back-and-forth between the two reasoning tasks.[5] Unfortunately, the instructions yield *no unique representation* of ST, and even if one considers only 4-cards settings, (NEC) is ambiguous, and induces *no unique ranking* (see S&L 2008, ch. 3 for details).

To get an idea of the difficulty of the task faced by subjects, it will be sufficient to restrict our attention to a 'generic' problem $\langle \mathbf{K}_{ST}, H_{ST}, Q_{ST} \rangle$ where: (*a*) every $\kappa \in \mathbf{K}_{ST}$ is characterized by four cards, with A, K, 4 and 7 as only possible values; (*b*) $h \in H_{ST}$ is generated by turning cards; and: (*c*) $Q_{ST}$ ={Rule, ¬Rule}, where Rule satisfies a material conditional. Other representations discussed will be *subsets of $H_{ST}$*.

Let us furthermore consider representations induced by *patient* and *credulous* LM, that (resp.) wait for 'Nature's answers' before they assess $Q_{ST}$, and do not doubt them. These two assumptions make reasoning purely 'deductive-interrogative' (with no ampliative reasoning, and as little redundancy as possible).[6] Even under such extremely favorable conditions, the ST problem is a tough nut to crack.

Under the above assumptions, the *ST problem* is a triple $P_{ST}=\langle \mathbf{K}_{ST}, H_{ST}, Q_{ST} \rangle$, where:

- $\mathbf{K}_{ST}$ ={{$(A, x_1), (K, x_2), (4, x_3), (7, x_4)$} : $x_1, x_2 \in \{4, 7\}$ and $x_3, x_4 \in \{A, K\}$}; and:
- For all $h \in H_{ST}$, $h(n) = \{(A, x_1), (K, x_2), (4, x_3), (7, x_4)\}$, with $x_1, x_2 \in \{\cdot, 4, 7\}$ and $x_3, x_4 \in \{\cdot, A, K\}$.[7] Furthermore every $h \in H_{ST}$ satisfies:

  **H0**  $h|1 = h_0 = \langle \{(A, \cdot), (K, \cdot), (4, \cdot), (7, \cdot)\} \rangle$ (back values are initially unknown); and

  **H1**  if $h(n) = \{(A, x_1), (K, x_2), (4, x_3), (7, x_4)\}$ and $e = \{(A, x'_1), (K, x'_2), (4, x'_3), (7, x'_4)\}$ then: $h \frown e \in H_{ST}$ iff: (*a*) $x_i \neq x'_i$, for some $x_i$ (one back is revealed at each position after the first) ; and: (*b*) if $x_i \neq \cdot$, then $x'_i = x_i$ (no value is forgotten once revealed).

- $Q_{ST}$ = {Rule, ¬Rule}, with $\kappa \in$ Rule iff $x_1 = 4$ and $x_4 = K$, or equivalently (given $\mathbf{K}_{ST}$): $\kappa \in$ ¬Rule iff $x_1 = 7$ or $x_4 = A$.

In any particular instance of $P_{ST}$, $\kappa_0$, the underlying *state of Nature* in $\mathbf{K}_{ST}$, is *uniquely identified at each maximal history*—the last position is identical with some $\kappa_i \in \mathbf{K}_{ST}$ s.t. $\kappa_i = \kappa_0$. Identifying $\kappa_0$ only 'up to' inclusion in Rule (or ¬Rule) suffices to assess $Q_{ST}$. Given that states which differ only w.r.t. unknown back values at $h(n)$, are indiscernible from $\kappa_0$ at $h(n)$, and that extensions of $h_0$ 'shrink' indiscernibility,[8] $\kappa_0$ need not be uniquely identified for $P_{ST}$ to be solved. Since $\mathbf{A}_{ST} = \emptyset$,

---

[5]Subjects may also rely massively on semantic 'precomputations' (see S&L 2008, § 4.2) for either or both tasks. OC's analysis may indicate that subjects 'precompute' the range of values on sides of cards, possibly from preconceptions about card decks (we thank Ingar Brinck for bringing this to our attention).

[6]Such LM treat the problem as one of *pure discovery* (see Hintikka et al. 1999), and are *Ockham efficient*, minimizing the number of retractions (see Kelly 2007).

[7]The first element of each pair stands for the initially visible side in some state $\kappa \in \mathbf{K}_{ST}$ or in $h$; and '·' indicates an unknown value at (the last position of) $h$.

[8]A formal definition would be: for every $h \in$ ST of length $n$, $h(n)$ induces an relation $R_h \subseteq \mathbf{K}^2_{ST}$, defined as:

any LM for $P_{ST}$ is a (partial) function $l : H_{ST} \longmapsto \mathbf{Q}_{ST} \times Q \cup \{?\}$, with (for any $l$ addressing $P_{ST}$). Without further restrictions:

$$\mathbf{Q}_{ST} = \{\texttt{turn}[S] : S \subseteq \{(A, \cdot), (K, \cdot), (4, \cdot), (7, \cdot)\}\} \qquad (\mathbf{Q}_{ST})$$

Note that a *complete* (extensional) representation of $\mathbf{K}_{ST}$ is not necessary to solve $P_{ST}$ if the rule is read as a material conditional. Intensional tests for the properties "being a state where $x_1 = 4$ and $x_4 = K$" or "being a state where $x_1 = 7$ or $x_4 = A$" are each sufficient to assess membership of equivalence classes that decide $Q_{ST}$ (intensional tests are cognitively more realistic than extensional ones, see S&L 2008, p. 178). While a (Nor)-like selection always fulfills *both* tests, a shorter selection may sometimes suffice, which expresses more formally as:

**Lemma 1.** *Selections of $(A, \cdot)$ and $(7, \cdot)$, together or in (any) sequence, are: (a)* sufficient *to solve* every *instance of $P_{ST}$; (b)* unnecessary *to solve* some *instances of $P_{ST}$.*

*Proof.* (*a*) Immediate from the definition of $Q_{ST}$. (*b*) $\texttt{turn}[(A, \cdot)]$ and $\texttt{turn}[(7, \cdot)]$ are sufficient when resp. $(A, 7) \in \kappa_0$ and $(7, A) \in \kappa_0$, to stabilize on ¬Rule, and decide $P_{ST}$. □

The importance of Lem. 1 comes from the ambiguity of (NEC), which can be given either *narrow* or *wide scope*. The former warrants only a 'history-bounded' heuristic use, in order to obtain, from some history $h$, a history $h^*$ identical with $h$ save for steps *redundant in $h$* to identify $\kappa_0$ up to inclusion in Rule or ¬Rule. The latter allows for 'cross-history' comparisons, i.e. to obtain, from some history $h$, a history $h^*$ identical with $h$ save for steps *redundant in some $h'$* (possibly distinct from $h$).

## 3.2   Solving the ST Problem

Let us call the *Wason* LM, denoted ($l_W$), the function defined in Fig. 1.; ($l_W$) decides $P_{ST}$ (see below), and the more constrained 'one-shot' ST problem $P^1_{ST}$, where admissible LM must stop after picking a *single element of $\mathbf{Q}_{ST}$*—i.e. where $\mathbf{K}^1_{ST} = \mathbf{K}_{ST}$, $Q^1_{ST} = Q_{ST}$, but $H^1_{ST}$ is generated by *one-time* selections.

$$l_W(h \frown e) = \begin{cases} \langle \texttt{turn}[(A, \cdot), (7\cdot)], ? \rangle & \text{if } h = h_0 \text{ and } e = \emptyset \\ \langle \texttt{turn}[\emptyset], \text{Rule} \rangle & \text{if } h = h_0 \text{ and } e = \{(A, 4), (K\cdot), (4\cdot), (7\cdot)\}, \\ \langle \texttt{turn}[\emptyset], \neg\text{Rule} \rangle & \text{if } h = h_0 \text{ and } e \neq \{\langle(A, 4), (K\cdot), (4\cdot), (7\cdot)\rangle\}, \\ \text{undefined} & \text{otherwise} \end{cases} \qquad (l_W)$$

Figure 1.

The ambiguity of (NEC) is irrelevant in $P^1_{ST}$: a card is redundant in some $h \in H^1_{ST}$ only if its selection reduces indiscernibility more than necessary in $h$ (narrow scope) or some possibly distinct

---

$R_h = \{\langle \kappa_i, \kappa_j \rangle : \kappa_i, \kappa_j \in \mathbf{K}_{ST} \text{ and } \kappa_i \cap \kappa_j = h(n) \setminus \{(x, y) : y = \cdot\}\}$. $R_h$ is an *equivalence relation* (reflexive, symmetric and transitive), and partitions $\mathbf{K}_{ST}$ between states excluded by, and compatible with, evidence available at $h(n)$.

$h' \in H^1_{\mathrm{ST}}$ (wide scope). But there is no way to sufficiently reduce indiscernibility in $P^1_{\mathrm{ST}}$ with a selection 'smaller' than $\mathtt{turn}[(A, \cdot), (7, \cdot)]$, unless one already knows that $\kappa_0 \in \neg\mathsf{Rule}$ *and* why (which value satisfies the intensional test of the 'disjunctive' property). This being blatantly circular, we have immediately:

**Observation 2.** $(l_W)$ *is the only* LM *deciding* $P^1_{ST}$ *under* (NEC).

Two immediate consequences of Obs. 2 are that: (*a*) $(l_W)$ is the only LM deciding $P_{\mathrm{ST}}$ with a *one-shot* selection without unnecessary cards; and: (*b*) to a competent subject whose representation of $P_{\mathrm{ST}}$ is limited to $P^1_{\mathrm{ST}}$, (Nor) is the *unique admissible selection* (and report). If the problem is viewed as a game against Nature, $(l_W)$ is a *uniform strategy*: its recommendation is identical, whatever the underlying state is (figuratively, whatever Nature's strategy is).

A uniform strategy or LM may be dominated depending on costs of questions, estimated probabilities of states and answers—and specifically in $P_{\mathrm{ST}}$, on the perceived value of exhaustive answers. There is no unique way to set these parameters, which instructions of § 2.1 leave unconstrained (see below for thematic versions). Fig. 2. displays two *non-uniform* strategies for $P_{\mathrm{ST}}$. When (NEC) is given narrow scope, it holds that:

**Observation 3.** *(a)* $(l_W)$ *is the only* uniform LM *deciding* $P_{ST}$ *without unnecessary cards. (b)* $(l_1)$ *and* $(l_2)$ *are the only* non-uniform LM *deciding* $P_{ST}$ *without unnecessary cards.*

$$
l_1(h \frown e) = \begin{cases}
\langle \mathtt{turn}[(A, \cdot)], ? \rangle & \text{if } h = h_0 \text{ and } e = \emptyset \\
\langle \mathtt{turn}[\emptyset], \neg\mathsf{Rule} \rangle & \text{if } h = h_0 \text{ and } e = e_1 = \langle (A, 7), (K, \cdot), (4, \cdot), (7, \cdot) \rangle, \\
\langle \mathtt{turn}[(7, \cdot)], ? \rangle & \text{if } h = h_0 \text{ and } e = e_2 = \langle (A, 4), (K, \cdot), (4, \cdot), (7, \cdot) \rangle, \\
\langle \mathtt{turn}[\emptyset], \neg\mathsf{Rule} \rangle & \text{if } h = h_0 \frown e_2 \text{ and } e = e_{2.1} = \langle (A, 4), (K, \cdot), (4, \cdot), (7, A) \rangle, \\
\langle \mathtt{turn}[\emptyset], \mathsf{Rule} \rangle & \text{if } h = h_0 \frown e_2 \text{ and } e = e_{2.2} = \langle (A, 4), (K, \cdot), (4, \cdot), (7, K) \rangle, \\
\text{undefined} & \text{otherwise}
\end{cases} \tag{$l_1$}
$$

$$
l_2(h \frown e) = \begin{cases}
\langle \mathtt{turn}[(7, \cdot)], ? \rangle & \text{if } h = h_0 \text{ and } e = \emptyset \\
\langle \mathtt{turn}[\emptyset], \neg\mathsf{Rule} \rangle & \text{if } h = h_0 \text{ and } e = e'_1 = \langle (A, \cdot), (K, \cdot), (4, \cdot), (7, A) \rangle, \\
\langle \mathtt{turn}[(A, \cdot)], ? \rangle & \text{if } h = h_0 \text{ and } e = e'_2 = \langle (A, \cdot), (K, \cdot), (4, \cdot), (7, K) \rangle, \\
\langle \mathtt{turn}[\emptyset], \neg\mathsf{Rule} \rangle & \text{if } h = h_0 \frown e'_2 \text{ and } e = e'_{2.1} = \langle (A, 7), (K, \cdot), (4, \cdot), (7, K) \rangle, \\
\langle \mathtt{turn}[\emptyset], \mathsf{Rule} \rangle & \text{if } h = h_0 \frown e'_2 \text{ and } e = e'_{2.2} = \langle (A, 4), (K, \cdot), (4, \cdot), (7, K) \rangle, \\
\text{undefined} & \text{otherwise}
\end{cases} \tag{$l_2$}
$$

Figure 2.

Obs. 3*a* follows from Obs. 2 and the above remarks. Obs. 3*b* follows from Lem. 1, and that using heuristically (NEC) with narrow scope does not warrant elimination of more possible moves than those already avoided by $(l_1)$ and $(l_2)$.

If a subject's representation of $P_{\mathrm{ST}}$ is rich enough to include either $l_1(H_{\mathrm{ST}})$ or $l_2(H_{\mathrm{ST}})$, some preference for uniform strategies is required to favor $(l_W)$. In $P_{\mathrm{ST}}$, uniform strategies exhaust potential counterinstances. This explains their being correlated with a strong preference for a rule not to be

violated (see O&C 1994).

Also, a subject who selects ($l_1$) or ($l_2$), will only 'tick' both ($A, \cdot$) and ($7, \cdot$) if she has some reason to report selection made in the 'longest' history, e.g. if she has a preference for shorter methods (or smaller representations) *and* perceives the game as *strictly competitive*—i.e. that Nature prefers a state which maximizes the number of moves necessary to decide $P_{ST}$—which seems unnatural (not to say flat-out irrational).

Finally, if a subject's representation includes $l_1(H_{ST})$ or $l_2(H_{ST})$, and her interpretation consider wide scope to (Nec) admissible, she will deem the problem possibly unsolvable, an evaluation that the proviso 'if any' may even facilitate. Indeed, if (Nec) is given wide scope, it follows immediately from Lem. 1*b* that:

**Observation 4.** *There is no* LM *deciding $P_{ST}$ without unnecessary cards.*

## 4 Conclusions

### 4.1 Do deontic aspects matter?

Wason's explicit assumption that (Nor) is normative, makes $P_{ST}^1$ the best candidate for being his own representation of the (abstract) problem (by Obs. 2). Coordinating on that representation is easy for readers who know Wason's favored solution. The 'logical' character of (Nor) has seldom been called into question, which shows Wason's assumptions influence precomputations of his learned readers. By contrast, relatively unschooled empirical subjects must, in order to recover $P_{ST}^1$, reason under a self-imposed restriction to one-shot selections. Otherwise (Nor)-like selection depend on particular preferences for uniform strategies, or expectations that Nature plays competitively.[9]

Variants with 'deontic' rules typically induce preferences for strategies that find *all* violators. (O&C 1994) exploits this feature to explain (Nor)-like selections in 'thematic' tasks, representing preferences with utilities, and using Bayesian decision theory. Our framework dispenses with computation (and maximization) of expected utility, and generalizes preference-based reasoning to abstract tasks. Adequately constrained preferences favor selections of uniform one-shot strategies, or sequential strategies that do not halt on success (and end up with the (Nor)-like selection in all histories), whether or not the rule is deontic.

Relevance theorists, who view logical and semantic reasoning as largely context-independent, have predicted that (Nor)-like selections would be elicited in contexts where deontic readings are excluded, through context-dependent biases induced by *pragmatic* 'relevance effects,' and validated that prediction empirically (see Girotto et al. 2001). While this model seems at odds with our

---

[9]S&L 2008, p. 105-106, 111-112 report dominant (Nor)-like selections in an competitive game, where subjects play against a program responding (unbeknown to them) to their selection so as to realize only the 'longest' histories. Shorter selections are favored, which makes the setting competitive, although subjects are unaware of this. S&L 2008 explain the dominance of (Nor) by the reduced computational cost of planning one 'run' (rather than multiple contingencies), but do not interpret this in game-theoretic terms, missing also the fact that the game is no longer vs. an unbiased Nature, making the task only remotely related to the original ST.

analysis, once understood that subjects can rely on 'precomputed' partial semantic representations, or implement automatically intensional heuristics, there is no obstacle to consider that contextual factors trigger 'default' *semantic* computations (see S&L 2008, pp. 113–114), nor to associate them with preference patterns.

A topic for future research is therefore to test the hypothesis that 'relevance effects' can be explained by constraints favoring (possibly precomputed) uniform strategies in abstract variants of the ST. If successful, this would offer a partial vindication of the *decision-theoretic* aspects of OC's rational analysis, initially (and incorrectly) restricted to 'thematic' tasks, through its generalization. The resulting model would outperform the explanatory power of Bayesian analysis, and overcome objections of Relevance theorists, without reverting to a classical competence model.

## 4.2 Heuristics without biases

Bayesian models impose unrealistic (intractable) computational demands on cognitive systems (see Binmore and Shin 1992, Kwisthout et al. 2011). Whether our model is, in that respect, an improvement, is an open problem. On the one hand, its treatment of semantic reasoning is related to *constraint satisfaction* in AI and logical programing, where a problem is identified with a set of variables, taking values from specific domains, and a set of constraints over possible value assignments to them. Heuristics for intractable problems are well studied, and may translate into LM and partial strategies.[10]

The role we give to preference- and heuristics-based reasoning to model bounded rationality, is influenced by evolutionary theories of inference, according to which logical reasoning exapted communication abilities (see Brinck and Gärdenfors 2003, Skyrms 2011).[11] This may seem at odds with Gigerenzer's well-known approach to bounded rationality through 'fast and frugal' heuristics (which also addresses the computational issue of Bayesian models), since he has claimed that heuristic-based 'ecological' rationality do not converge through evolution towards 'normative' standards of logical reasoning and decision-making (see e.g. Gigerenzer 2000). However, iterations of 'fast and frugal' choice heuristics converge to well-behaved choice functions (see Arlò-Costa and Pedersen 2011). And LM *qua* heuristics are partial (algorithmic) strategies, sometimes sufficient to solve classical underlying games with tractable reasoning (see Halpern and Rego 2006). Empirical results in ST can be explained by the failure of experimenter to constrain the representation of the game enough for the subject to choose the appropriate solution, yielding a uniform (game-theoretic) model for both reasoning 'to' and 'from' an interpretation in problem-solving tasks.

---

[10]We thank Claes Strannegård for pointing to us the connection with constraint satisfaction.

[11]In particular, we have argued in (Genot and Jacot 2012b) that consequence relations (classical or otherwise) can supervene on heuristics for treatment of semantic information in argumentative games.

# References

Arlò-Costa, H. and Pedersen, A.P. (2011). Bounded Rationality: Models for Some Fast and Frugal Heuristics. In van Benthem *et al.* (eds), *Games, Norms and Reasons*, pp. 1–21. Springer Netherlands.

Binmore, K. and Shin, H.S. (1992). Algorithmic Knowledge and Game Theory. In Bicchieri and Della Chiara (eds), *Knowledge, Belief and Strategic Interaction*, ch. 9, pp. 141–154. Cambridge University Press.

Brinck, I. and Gärdenfors, P. (2003). Cooperation and Communication in Apes and Humans. *Mind & Language*, 18(5):484–501.

Genot, E.J. and Jacot, J. (2012b). Semantic Games for Algorithmic Players. 6th Workshop on Decision, Games and Logic, June 2012, Munich (Germany).

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World*. Oxford University Press.

Girotto, V., Kemmelmeier, M., Sperber, D., and van der Henst, J.-B. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition*, 81(2):B69–B76.

Halpern, J.Y. and Rego, L.C. (2006). Extensive Games With Possibly Unaware Players. In *Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 744–751. AAMAS.

Hintikka, J., Halonen, I., and Mutanen, A. (1999). Interrogative Logic as a General Theory of Reasoning. In *Inquiry as Inquiry*, pp. 47–90. Kluwer.

Kelly, K. T. (2007). Simplicity, Truth, and the Unending Game of Science. In Bold *et al.* (eds), *Foundations of the Formal Sciences V: Infinite Games*, pp. 223–270. College Publications.

Kwisthout, J., Wareham, T., and van Rooij, I. (2011). Bayesian Intractability Is Not an Ailment That Approximation Can Cure. *Cognitive Science*, 35(5):779–784.

Martin, E. and Osherson, D. (1998). *Elements of Scientific Inquiry*. The MIT Press.

Oaksford, M. and Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101(4):608–631.

Skyrms, B. (2011). Pragmatics, Logic and Information Processing. In Benz *et al.* (eds) *Language, Games, and Evolution*, pp. 177–187. Springer Berlin / Heidelberg.

Stenning, K. and van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. The MIT Press.

Wason, P.C. (1968). Reasoning about a rule. *Quaterly Journal of Experimental Psychology*, 20(3):273–281.