Semantic Games for Algorithmic Players

Genot, Emmanuel; Jacot, Justine

2012

Link to publication

Citation for published version (APA):

General rights
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

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.
Semantic Games for Algorithmic Players

Emmanuel J. Genot · Justine Jacot

Abstract We describe a class of semantic extensive entailment game (EEG) with algorithmic players, related to game-theoretic semantics (GTS), and generalized to classical first-order semantic entailment. Players have preferences for parsimonious spending of computational resources, and compute partial strategies, under qualitative uncertainty about future histories. We prove the existence of local preferences for moves, and strategic fixpoints, that allow to map EEG game-tree to the building rules and closure rules of Smullyan’s semantic tableaux (ST). We also exhibit a strategy profile that solves the fixpoint selection problem, and can be mapped to systematic constructions of semantic trees, yielding a completeness result by translation. We conclude on possible generalizations of our games.

1 Introduction

For some first-order language $L$, $\Gamma \subseteq L$ and $\phi \in L$, $\Gamma$ semantically entails $\phi$ (noted $\Gamma \models \phi$) iff all models of $\Gamma$ are also models of $\phi$, noted $\text{Mod}(\Gamma) \subseteq \text{Mod}(\phi)$. Proving that entailment holds can be modeled as a two-player extensive entailment game (hereafter EEG), where Abelard chooses a model for $\Gamma$, and Eloise must then show it to be a model of $\phi$. If she has a winning strategy (w.s.) exactly when $\Gamma \models \phi$, winning conditions for Eloise coincide with semantic entailment.

Choice of a model can be substituted with selection of truth conditions for $\Gamma$ and $\phi$, and models are ‘read off’ from branches. A game where Eloise has a w.s. exactly when each of Abelard’s possible selection of truth conditions either satisfies $\phi$ or is inconsistent, captures equally entailment. If two structures $\mathcal{M}_1$ and $\mathcal{M}_2$ are isomorphic (noted $\mathcal{M}_1 \cong \mathcal{M}_2$), and if two interpretations of $L$ $I_{\mathcal{M}_1}$ and $I_{\mathcal{M}_2}$ preserve isomorphism, then $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfy the same formulas (Manzano, 1999, Th. 2.46); in particular they have the same complete...
diagram (description through atoms of $\mathcal{L}$ and their negations). Hence, a semantic entailment game where each run concludes with each player being committed to a complete diagram, offers an in-game syntactic test for comparison of models.

More formally, an attempt to prove $\Gamma \models \phi$ can be modeled as an extensive entailment game between Abelard ($A$) and Eloise ($E$), where: $A$ ($E$) can ask $E$ ($A$) to commit to subformulas of $\phi$ (resp.: of some $\gamma \in \Gamma$); and $A$'s ($E$'s) commitments can (in the limit) yield a complete diagrams for some $A \in \text{Mod}(\Gamma)$ ($E \in \text{Mod}(\phi)$). $E$ must show that $A \in \text{Mod}(\phi)$—or, more generally, that there is a $E \in \text{Mod}(\phi)$ such that $A \not\simeq E$. Therefore, a game where $E$ has a w.s. exactly whenever she can match $A$’s possible selections of literal either satisfies $\phi$, or is inconsistent, captures entailment from $\Gamma$ to $\phi$.

In this paper, we describe a class of semantic extensive entailment game, related to game-theoretic semantics that capture first-order classical first-order semantic entailment in the above sense. Players of these games are algorithmic: they compute only partial representations of the game, and partial strategies. In section 2, we first motivates our algorithmic approach (2.1), then define the game (2.2) and the strategic preferences of algorithmic players (2.3); from the latter two, local preferences for moves are obtained (2.4).

Section 3 details players’ reasoning: we define strategic fixpoints that both players aim at (3.1); and their best responses when reaching them (3.2). From this, we obtain a translation scheme w.r.t. signed semantic trees (3.3) and exhibit a strategic profile that generates a closed game-trees iff $E$ has a w.s., which entails their completeness for classical first-order entailment (3.4).

Section 4 discusses extensions to model abilities of semantically sophisticated players (4.1), connections with standard game-theoretic semantics (4.2), and learning-theoretic aspects of our games, and their extension to other consequence relations (4.3). We conclude on the relation between our model, game-theoretic models of communication, and the problem of understanding human linguistic and logical competence.

## 2 Semantic entailment games

### 2.1 Motivations

When logical consequence can be analyzed with games, logical reasoning can be accounted for as reasoning about games. In particular, one can employ the method of epistemic game theory, which studies how players solve games reasoning from assumptions about the game setting (in particular, other players’ preferences). Classical EGT explains strategy selection as a ‘top-down’ process, from a complete representation of the game. Therefore, classical
EGT applies to entailment games only if players have a representation of all possible models \( \Gamma \) and \( \phi \)—or all possible ways to obtain their diagrams.

*Game-theoretic semantics* (GTS) and *dialogical semantics* (DS), as exposed resp. by Hintikka and Sandu (1997) and Rahman and Keiff (2005), both study semantic games where attacks and defenses are based on semantic clauses. These games capture *material* truth (GTS) or *logical* truth (DS), through existence of w.s. in extensive games. Both proceed ‘bottom up,’ explaining players’ strategies as gradual analysis of truth conditions, and impose restrictions on players’ strategies, so that DS games yield game trees that correspond to *semantic trees* (ST) developed by Smullyan (1968) as do GTS games, when extended to entailment (see e.g. Harris, 1994).

In ST, branches generated by semantic clauses select truth conditions verifying \( \Gamma \) and falsifying \( \phi \). When a branch reaches a contradictory assignment, the branch is closed after finitely many steps. Systematic constructions yield trees with finitely many closed branches iff \( \Gamma \) entails \( \phi \). Both DS and (extended) GTS explain strategically selection of options that generate trees similar to those systematic constructions of ST through strategic considerations, without appeal to a complete representation of the game. However, in the absence of well-defined preferences, restrictions to recover closure rules, or obtain finite games whenever possible, etc. remain *ad hoc*.

A systematic correspondence between EEG and ST must be grounded in preferences of players and explain their ‘bottom-up’ reasoning. It has to map *symmetric* semantic clauses of EEG to *asymmetric* tree-building rules of ST, as well as closure rules of the latter, to strategic fixpoints in the former, by *in-game inferences*, in the fashion of EGT. Finally, it must exhibit a strategy profile (a pair of strategies) that solve the fixpoints selection problem.

### 2.2 Definition

Following Halpern and Rego (2006), partial anticipations can be modeled with (player-indexed) awareness function mapping positions to representations of the game. Partial strategies are partial functions from the set of positions the game may reach, recommending actions for positions they consider. Players will typically need several runs to assess entailment: outcomes of pairs of strategies will be *game trees* rather than single history. The definition of the EEG is as follows:

**Definition 1.** An EEG is a pair: \( \mathcal{G}(\Gamma, \phi) = (\mathcal{U}(\Gamma, \phi), (A_i)) \) where \( \Gamma \subseteq \mathcal{L} \) and \( \phi \in \mathcal{L} \); and:
\[ \begin{align*} &X\psi_1 \lor \psi_2 \\
 &\downarrow \\
 &Y(\psi_1 \lor \psi_2, \emptyset) \\
 &\downarrow \\
 &X\psi_1 \quad X\psi_2 \\
 &\downarrow \\
 &X\psi_1 \quad X\psi_2 \\
 &\downarrow \\
 &X\neg \psi_1 \quad X\psi_2 \\
 &\downarrow \\
 &Y(\neg \psi_1) \\
 &\downarrow \\
 &X\neg \psi_1 \\
 &\downarrow \\
 &Y(\neg \psi_1, \neg \psi_2) \\
 &\downarrow \\
 &X\neg \psi_1 \quad X\psi_1 \\
 &\downarrow \\
 &X\neg \psi_1 \quad X\neg \psi_1 \\
 &\downarrow \\
 &X\neg \psi \quad X\neg \psi \quad X\neg \psi \\
 &\downarrow \\
 &Y(\neg \psi, \psi) \\
 &\downarrow \\
 &X\psi \\
 &\downarrow \\
 &X\neg \psi \\
 &\downarrow \\
 &X\exists x\psi(x) \\
 &\downarrow \\
 &Y(\exists x\psi(x), \emptyset) \\
 &\downarrow \\
 &X\exists x\psi(x) \\
 &\downarrow \\
 &Y(\exists x\psi(x), \psi) \\
 &\downarrow \\
 &X\exists x\psi(x) \\
 &\downarrow \\
 &\{\psi \subseteq \{\psi(a_j) : j \in \mathbb{N}\} \} \\
 &\downarrow \\
 &X\psi \\
 &\downarrow \\
 &\{\psi \subseteq \{\psi(a_j) : j \in \mathbb{N}\} \} \\
 \end{align*} \]

Fig. 1

- \(U(\Gamma, \phi) = (N, H, P, (z_i))\) is the underlying classical game, where: (a) \(N = \{A, E\}\) is the set of players; (b) \(H\) is a set of histories, i.e. sequences of actions, with \(Z \subset H\) including terminal histories—with either \(X\text{-stop}\) as last position of \(h\), or infinite; (c) \(P : H \to N\) is a player function s.t. if \(P(h) = X\), then \(X\) moves after \(h\); (d) \((z_i) = \{z_A, z_E\}\) are preferences over \(Z\).

- \((A_i) = \{A_A, A_E\}\) are awareness functions s.t. \(A_X : H \to 2^H\) and for all \(h, h', \text{ if } h = (h'|m), A_X(h) \subseteq A_X(h')\) (perfect recall).

We denote \(M_X(h)\) the actions available to \(X\) at \(h\) \((M_X(h) = \emptyset\) whenever \(P(h) = Y\)). For every \(h\) s.t. \(P(h) = X\), \(M_X(h)\) includes analytic and nonanalytic queries. By the first, based on semantic clauses (displayed Fig. 1), \(X\) can \(Y\) to commit to subformulas of \(Y\)'s former statements. The second, denoted \(X?(\psi \lor \neg \psi)\), is equivalent to addressing a 'yes-no' question to \(Y\) (introducing its presupposition). Given classical meanings of connectives, \(Y\) must eventually answer—but may delay the answer indefinitely. Following Hintikka (1986), we call the set of yes-no questions \(X\) is ready to ask in a EGG the range of attention of \(X\).

Players can in principle play until complete diagrams are obtained. With \(h, h' \in Z\) two such histories, and assuming a map sending \(h\) and \(h'\) to pairs \((A, \mathcal{E})\) and \((A', \mathcal{E}')\), where \(\mathfrak{A}, \mathfrak{A}' \in (\Gamma)\) and \(\mathcal{E}, \mathcal{E}' \in (\phi)\), \((z_i)\) capture classical entailment if whenever \(h\) and \(h'\) are mapped (resp.) to \((\mathfrak{A}, \mathcal{E})\) and \((\mathfrak{A}', \mathcal{E}')\), then if \(\mathfrak{A} \subseteq \mathcal{E} \text{ and } \mathfrak{A}' \not\subseteq \mathcal{E}'\) then \(h \Rightarrow_{\mathcal{E}} h'\) and \(h' \Rightarrow_{\mathfrak{A}} h\). Partial representation and anticipations prevent elimination of strategies that lead to inconsistent diagrams. However, an inconsistent diagram resulting from \(X\)'s choices can be mapped to a

---

\(^1\) Technically, \(H\) is s.t. the empty sequence is in \(H\); if \(h \in H\) is of length \(n\), then for all \(m \leq n\), \((h|m)\), the initial segment of length \(m\) of \(h\), is in \(H\) too; and an infinite \(h^\omega\) is in \(H\) if \((h|m) \in H\) for every finite \(m\).
‘pseudo structure’ $\mathcal{X}_i$, and the definition of $\varphi_X$ extended as follows: for any two $\mathcal{X}$ and $\mathcal{Y}$, if $h$ and $h'$ are mapped (resp.) to $(\mathcal{X}_i, \mathcal{Y}_i)$ and $(\mathcal{X}_i', \mathcal{Y}_i')$, $h \Rightarrow_{\mathcal{Y}} h'$ and $h' \Rightarrow_{\mathcal{X}} h$; and if $h$ and $h'$ are mapped (resp.) to $(\mathcal{X}_i, \mathcal{Y}_i)$ and $(\mathcal{X}_i', \mathcal{Y}_i')$, $h \sim_{\mathcal{Y}} h'$ and $h' \sim_{\mathcal{X}} h$. Finally, if $\mathcal{X}$ plays $\mathcal{X}$-stop, $\mathcal{Y}$ loses by default: Formally, if the last position of $h$ is $\mathcal{X}$-stop then $h \Rightarrow_{\mathcal{Y}} h'$ and $h' \Rightarrow_{\mathcal{X}} h$ for any terminal history $h'$.

2.3 Strategic Preferences

Our aim being to model players whose use of computational resources (when e.g. computing strategies) is as limited as possible, we need some assumptions about their limitations. And since EEG are games, we need also assumptions as to what is common knowledge between players in them.

**Assumption 1.** Between two (partial) strategies leading to the same expected result w.r.t. $\varphi_A$ or $\varphi_E$, both A and E will favor the one with fewer moves than the other.

**Assumption 2.** Both players understand that: (a) the outcome of each run of the game should be decided by a comparison of models w.r.t. their preferences; (b) in order to have a w.s., A needs to win only one run, while E has to win all possible runs.

**Assumption 3.** Both A and E give logical constants their classical first-order meaning.

**Assumption 4.** Player’s ranges of attention are not in general common knowledge.

**Ass. 1** makes moves akin to steps in a program, requiring resources to run on hardware. It should be read *ceteris paribus*, because redundancies may prevent ‘back-tracking’ and reduce the overall cost of storing and accessing a representation, and **Ass. 1** may therefore be locally violated. The assumption embodies a reasonable understanding of quantitative ‘resource-consciousness’ (spending as little as possible), while the *ceteris paribus* allows for qualitative modulation. As such, it is justified in the context of EEG.

**Ass. 2** and **Ass. 3** guarantee the classical interpretation of the consequence relation in the game. Since A and E may be unable to carry a comparison in finite time, they will need to rely on estimates. Together, they define implicitly the ‘semantic competence’ required to play for classical entailment.

**Ass. 4** stems from the fact that assuming classical meaning (and the excluded middle) nonanalytic queries are equivalent *Cut rule*, which can shorten proofs (see Boolos, 1984), whose ‘best’ use is incomputable. Therefore, algorithmic players (who compute strategies) cannot anticipate each other’s questions in general.

Because of **Ass. 1**, players should expect each other to attempt to use shortcuts, but without possibility to anticipate them, the following ‘strategic principle’ will apply:
S.P. 1 (From Harsanyi (1977)). If $X$ cannot rationally expect $Y$ to play any strategy other than the most harmful for $X$, then $X$ should play the strategy that is $X$’s best response to $Y$’s most harmful strategy.

Ass. 2 and 3 suffice to $A$ and $E$ to understand each others’ preferences and understand what the ‘most harmful’ strategy, and the best response to it, are. Assuming that it is common knowledge that players follow S.P. 1, is tantamount to assume them to be competent players. Given Ass. 2–3 players can also be assumed to obey the following principle (where $(-)Pã$ is a literal $Pã$ or $¬Pã$):

S.P. 2. $E$ should not reply to any query with $E(¬)Pã$, unless she can also use a query (or sequence thereof) to obtain $A(¬)Pã$. Equivalently: $A$ should compel $E$ to reply $E(¬)Pã$ to some query, for some $(¬)Pã$ s.t. she cannot obtain from $A$.

Diagram identity provides a syntactic of test for isomorphism, provided that players agree on how to apply nonlogical vocabulary to individuals (we will make this assumption in what follows). The best way for $E$ to enforce S.P. 2, is to never state a literal unless $A$ has stated it first; and conversely, the best way for $A$ to comply with it, is to state new literals whenever possible. We will take it to be the strategic content of S.P. 2 for $E$ and $A$, respectively.

Finally, provided that the prospect of loosing is never infinite, it follows from Ass. 1 that:

S.P. 3. If for some $h ∈ H$ s.t. $P(h) = X$, (a) $X$ can choose between two strategies $s$ and $s′$ s.t. $s$ and $s′$ resp. extend $h$ finitely and indefinitely; and: (b) $X$ cannot expect to be better off following $s′$ rather than $s$; then $X$ should prefer $s$.

S.P. 3 is justified if $A$ and $E$ are realized by agents who play multiple EEG (either in sequence, or in parallel), and prefer not to get stuck playing indefinitely in one game.

2.4 Preferences over moves

Ass. 2, together with S.P. 2, induce an asymmetry between players, $A$’s choices determine gradually a diagram, while $E$ merely has to match $A$’s choices. This asymmetry defines a partial preference ordering for $A$ and $E$ at their turns, which yields the following (proof in Appendix A).

Observation 2. Ass. 1–4 and S.P. 1–3 induce preferences for queries and replies as displayed Fig. 2

As a consequence of Obs. 2, the game as played by algorithmic players is finite, although the underlying game is not in general (when $L$ has countably many names).

---

2 $P$ is a $n$-ary predicate, and $ã$ a sequence of $n$ individual terms
3 Epistemic Game Theory for eeg

3.1 Socratic positions and fixpoints

Let us define Socratic positions—after Socrates, who used only concessions from his opponents as arguments—abbreviated Σ.P. (to avoid the confusion with S.P.), as follows:

**Definition 3.** A Socratic Position in some \( h \in H \) is a position \( m \) s.t.: (a) at \( m' < m, A \) has already targeted every \( E \)-labeled statement at least once; and: (b) at \( m, E \) has answered all of \( A \)'s queries complying with S.P. 2.

If Σ.P. are stable over extensions, \( E \) is guaranteed to always match \( A \)'s choices of literals, until a diagram is obtained. The following shows that they are:

**Observation 4.** For any history \( h \) of \( \mathcal{G}(\Gamma, \phi) \) s.t. \( P(h) = A \) and the last position \( m \) of \( h \) is a Σ.P., there is a recursive strategy for \( E \) such that, if \( A \) extends \( h \), and \( E \) follows that strategy, if \( E \) is eventually committed to a full diagram in some extension \( h' \) of \( h \), then it is identical to a full diagram \( A \) is also committed to in \( h' \).

(Proof in appendix.) In any given run, \( E \) is aiming at reaching a Σ.P., since it allows her to assess with certainty that her strategy will be successful in that run. Once the first Σ.P. of the run has been reached, Obs. 4 allows her to base her current estimate, and any future one, on the configuration at that position. Socratic positions are therefore strategic fixpoints.\(^3\)

\(^3\) This can be explicited defining partial estimate functions \( \text{Est}_X \) indexed on players, sending each position \( m \) of some history \( h \) to a position \( m' \in h' \), where either \( h' = h \) or \( h = h|m'' \) for some \( m'' \), such that \( \text{Est}_X(m) \) (when defined) is the closest position from \( h|0 \) where \( X \) can estimate who will win the current run. A Σ.P. \( m \) is therefore always mapped to itself by that type of function. More simply, but also more indirectly, is also a fixpoint for a learning function parasitic on the game discussed § 4.3.
cannot alone force the game to reach such a $\Sigma.P.$—$A$ can in principle prevent the game to reach any. The next section shows that $A$’s best interest is also to aim for $\Sigma.P.$, albeit for different reasons as $E$’s.

3.2 Reasoning from fixpoints

By Ass. 2b, players understand that $A$’s loosing a run is no indication that $E$ has a w.s. for the game. Halting a run that $A$ would loose, as soon as possible, and move to another, is therefore a sound strategy for $A$ that additionally complies with Ass. 1. The connection with $\Sigma.P.$ is as follows:

**Observation 5.** If, for some history $h$ of $G(\Gamma, \phi)$ such that $P(h) = A$, the last position $m$ of $h$ is a $\Sigma.P.$; and if the prospect of loosing a run of $G(\Gamma, \phi)$ is not infinitely negative; then $A$’s best strategy is to extend $h$ to $h' = (h, A$-stop).

(Proof in appendix.) $A$ may be lead to state an inconsistent diagram—$E$ will not, unless $A$ has, because of S.P. 2. Moreover $E$ can always avoid being asked a contradiction by restating some reply in $h$, unless $\phi$ has no models.4

Positions at which $A$ do so are also fixpoints, as established by the following:

**Observation 6.** If, for some history $h$ of $G(\Gamma, \phi)$ such that $P(h) = A$ and: (a) $P\bar{a}$—or $\neg P\bar{a}$—occurs in $h$; and: (b) $E(\psi, \neg P\bar{a})$—resp. $E(\psi, P\bar{a})$—also occurs at the last position of $h$; then $A$’s best option is to extend $h$ in $h' = (h, A$-stop).

(Proof in appendix.) Obs. 4 shows what $E$’s (local) strategy should be when a $\Sigma.P.$ is reached, and that planning it depends only on the past history. Obs. 5 and 6 show how $A$’s (local) reasoning determines his strategy when $\Sigma.P.$ are encountered.

3.3 A Translation Scheme

Partial strategies of players of an eeg generate gradually a partial representation of $U(\Gamma, \phi)$—or, equivalently, a subset of $H$. Obs. 2 shows a local topology that is reminiscent of signed semantic trees (st) of Smullyan (1968), where the premises and conclusion are prefixed (resp.) with $T$ (for true) and $F$ (for false), and where the tree formalizes the attempt to

4 If in some $h$, $E$ cannot adjust past choices to avoid being demanded a pair of inconsistent literals, then $A$ can always force such a position in $h$. Since only $A$’s moves generate histories, $A$ can also always force the game to $h$, and (eventually) any diagram for $E$ to be inconsistent. Hence, ther no set of consistent literal is compatible with $\phi$, and $\phi$ has no model. Contraposin, if $\phi$ is consistent, $E$ can adjust past her choices to avoid being demanded a contradiction.
obtain a proof by *reductio* that the premises entail the conclusion. **Fig. 1** translates in ST-*building rules* by **Obs. 2** if: (a) A and E are substituted with T and F, respectively; and (b) queries are omitted. Nonanalytic queries are equivalent to the *Cut* rule in tableau proof (see Boolos, 1984).

Given **Obs. 5** and 6, **A-stop** is equivalent of *closure rules* of ST for the case where (resp.) both T(-)Pã and F(-)Pã occur in the same branch, or both TPã and T¬Pã do. Since **Ass. 2** requires E to win every play where *diagrams* are compared, one can allow E to play **E-stop** when asked a contradiction in some history h (indicating that she mus backtrack her choices). Although h is ‘neutralized’ for the application of **Ass. 2**, this allows to complete the correspondence with st: **E-stop** is then equivalent to the closure that applies when both FPã and F¬Pã occur in the same branch. Therefore, every game-tree for some EEG can be mapped to some ST (with the *Cut* rule, if nonanalytic moves are played).

Conversely, for every ST T: (a) T can be extended to ST T′ where rules are applied until literals are reached; and: (b) T1 can be a ‘pruned’ into T″ where no building rule is applied after the first application of a closure rule, without effect for the assessment of whether Γ ⊨ φ; and: (c) omitting F-labeled nodes that would correspond to E-labeled moves not complying with **S.P. 2**, without effect for the assessment of whether Γ ⊨ φ either (see (Rahman and Keiff, 2005, § 2.5); their result assumes a restriction equivalent **S.P. 2**, although not derived from players’ preferences). Hence, every ST T is (partially) realized by an EEG game-tree (with only analytic moves, if the ST is *Cut-free*).\(^5\)

The above argument can be summed up in the following lemma:

**Lemma 7.** Given **Ass. 1–4** and **S.P. 1–3**, there is a one-one correspondence between local strategies for selection moves in extensive entailment games, and building and closure rules of signed semantic trees.

3.4 Existence of a solution

By **Lem. 7**, exhibiting a strategy profile that solves the Σ.P. selection problem whenever Γ entails φ is sufficient to prove that EEG capture first-order entailment, because it amounts to solving a building scheme for ST that closes whenever Γ ⊨ φ. A simple solution is to ‘rear-engineer’ such a profile from systematic constructions of ST used to prove their completeness for first order entailment, which is exactly characterized by the existence of a closed ST with

\(^5\) The realization is partial, because branches of an EEG are built sequentially, while ST are built in parallel: an ST with one open infinite branch can be realized by an EEG game-tree with only one (infinite) branch, if the infinite branch is the first to be explored by the players. However, a ST is realize by (the union of) a family of EEG.
premises $\Gamma$ and conclusion $\phi$. If this is possible, Lem. 7 guarantees that the profile will match A’s and E's local preferences for moves.

Reaching a $\Sigma.P.$ whenever possible, and as quickly as possible, in each run, is in both players interest (see § 3.2). For A, it prevents long explorations of run he would eventually lose, and he can let E play (e.g. repetitions or nonanalytic moves) if none is reached. For E, it makes for faster exploration of histories, because A’s best response is to move to explore another branch.

Let $O(s_a, s_e)$ denote the the game-tree explored when A and E implement (resp.) strategies $s_a$ and $s_e$. A strategy profile can be obtained, through Lem. 7, from systematic $st$, defining a pair of systematic strategies, which are partial, and can be expressed as sets of instructions for using queries and reply (Q/R), halting (H), and ‘switching’ histories (S) when playing a new run is necessary.6 A systematic strategy $s^*_a$ for A in $G(\Gamma, \phi)$ is any (partial) strategy for A, s.t.:

Q/R (a) $s^*_a$ uses analytic queries against $\phi$ (and its subformulas), until queries for literals are exhausted (without repetition), and in case of ‘branching’ explores the leftmost history first; (b) $s^*_a$ replies without delay, and in case of ‘branching’ explores the leftmost history first; (c) $s^*_a$ uses nonanalytic queries to obtain a complete diagram only, and never introduces new individuals that way.

H $s^*_a$ selects A-stop as soon as: (a) a $\Sigma.P.$ is reached; or (b) a query from E asks A for a statement that contradicts a previous (A-labeled) one;

S after any terminal history $h$, $s^*_a$ repeats (if possible) the same offensive sequence as in $h$ until the lowest ‘branching’ with an unexplored branch in $h$ is reached, and then explore it.

Guidelines (Q/R) aim at reaching a position of comparison as quickly as possible; (H) follows to the letter the ‘best response’ of Obs. 5 and Obs. 6, and (S) is a ‘backtracking’ scheme to explore systematically alternative histories, if possible (i.e. depending on E’s moves). In short, $s^*_a$ simply adds to Ass. 1–4, and S.P. 1–3, a scheme for systematic exploration of the games, that will be successful if E follows some systematic strategy as well.

A good systematic strategy for E lets A target $\phi$ to know which literal(s) are to be obtained, then works towards obtaining it (them) as quickly as possible. However, this may requires insights about the premises, or some estimate of possible lemmas (see § 4.1). The

---

6 When several moves are compatible with instructions of X’s strategy, X can be for all practical purposes thought to choose at random within admissible moves (i.e. to adopt a behavioral extension of the partial strategy).
following strategy $s^*_e$ for $E$ in $G(I, \phi)$ presupposes no such insight, and is advisable for a clueless $E$. It includes some redundancies, that are discussed after the definition.

Q/R (a) $s^*_e$ lets $A$ target $\phi$, until a literal is requested; (b) $s^*_e$ replies to analytic queries as soon as possible, an makes adjustments later, unless the reply is a literal; (c) if $A$ uses a nonanalytic query, $s^*_e$ recommends a ‘copycat’ strategy; and (d) once a literal is asked, it proceeds (recursively) as follows:

Initial Stage $s^*_e$ orders the premises, and targets the first premise $\gamma_1 \in I$ in the ordering;

this completes the first stage;

Stage $n$ Assume that the stage $n-1$ has been completed; either all the needed literals are obtained, or not; then:

1. if a $\Sigma.P.$ has been reached, then $s^*_e$ picks an arbitrary query, and repeats it until $A$ plays $A\text{-stop}$;

2. if not, in the current history $h$, $s^*_e$ targets the highest position $m$ featuring some $A\psi$ not yet targeted; then:

(a) if $A\psi$ is not of the form $A\neg\exists x\psi(x)$, $s^*_e$ plays $E(A\psi, \Psi)$, where $\Psi$ is chosen (when nonempty) according to preferences in Fig. 2;

(b) if $A\psi = A\neg\exists x\psi(x)$, then: (i) $s^*_e$ picks the constant $k_i$ occurring at the ‘highest’ earlier position in $h$ (including as an argument in a query), s.t. $A\neg\psi(x/k_i)$ does not occur in $h$ prior to $m$; and $s^*_e$ plays $E(\neg\exists x\psi(x), \psi(x/k_i))$; and: (ii) at $E$’s next move, $s^*_e$ plays $E'(A\neg\exists x\psi(x) \lor \neg A\neg\exists x\psi(x))$.

3. At $E$’s next move, $s^*_e$ plays $E?((\gamma_n \lor \neg \gamma_n)$, and never targets $m$ again. This completes the $n$th stage.

H $s^*_e$ selects $E\text{-stop}$ iff a query from $A$ asks $E$ for a statement that contradicts a previous ($E$-labeled) one.

Redundancies in $s^*_e$ result from the systematic inductive scheme of (Q/Rd), which may ask for literals $E$ may not need (and multiple occurrences); and uses redundant nonanalytic queries to make ‘backtracking’ easier. Although these features mimic systematic ST, they can be optimal for some player types: the first, as a way to obtain systematically literals with constants introduced in $h$, with a simple search algorithm; and the second, as a way to reduce working memory load (when ‘scanning’ upward for possible moves).

The strategy profile $O(s^*_A, s^*_e)$ is therefore well-motivated when: (i) $A$ lacks insights to select countermodels to $\phi$; and: (ii) $E$ lacks insights on how to obtain literals requested from her. Both are systematic, if unsophisticated, schemes to demand as many literals as possible,

7 Explicitly, if $A$ uses $A?\psi \lor \neg \psi$ at $m$ $s^*_e$ plays $E?\psi \lor \neg \psi$ at $m+1$, and waits for $A$’s reply, and copies it if it occurs, and leaves the query without answer if it does not.
that still avoid extension past fixpoints. Notice that E is led to contradict herself, this will always occur prior to \( s^*_E \) takes the offensive. Hence, once \( s^*_E \) has take the offensive, it will play until a \( \Sigma \text{.P.} \) is reached, or runs out of A-labeled moves to target. The outcome of \( O(s^*_a, s^*_E) \) is a ‘sequential’ version of the systematic stT called completed tableaux (see Smullyan, 1968, p. 63). It generates a finite game with finite horizon when \( \Gamma \models \phi \), or when all ‘open’ branches are finite; or a finite game with one infinite history when the first infinite ‘open’ branch is reached. The proof is a technical exercise (left to the reader) which yields a completeness result by translation. A game-theoretic formulation of this completeness result is:

**Observation 8.** For any EEG \( G(\Gamma, \phi) \), there is a strategy profile \((s^*_a, s^*_E)\), where \( s^*_a \) and \( s^*_E \) are (resp.) A’s and E’s strategies, s.t. \( s^*_E \) is a winning strategy for E iff \( \Gamma \models \phi \).

### 4 Discussion

#### 4.1 Sophisticated players

Although both \( s^*_a \) and \( s^*_E \) give guidelines to use nonanalytic moves, none is played in \( O(s^*_a, s^*_E) \). Nonanalytic queries are equivalent to the Cut rule, and it is well-known since (Boolos, 1984) that cut-free proofs can be dramatically long. If E anticipates that some \( \psi \) follows from \( \Gamma \), and that \( \{ \psi \} \) entails \( \phi \), she can introduce as her first move \( E? (\psi \lor \neg \psi) \). If she is correct, and A replies with \( A\psi \), she will be to win the (shorter) subgame \( G(\{ \psi \}, \phi) \); and if he replies with \( A\neg \psi \), she can force him to a contradiction, and to eventually explore histories where \( \psi \) holds. Without nonanalytic queries, i.e. without ‘cut’ or lemmas, there is

---

8 The proof in Smullyan (1968) shows that there exists at least one completed tableau, for any pair \( \{ \Gamma, \phi \} \) (Theorem 5, p. 64; from this, one obtains that a systematic tableau \( T(\Gamma, \phi) \) closes iff \( \Gamma \) entails \( \phi \), from the proof that \( T(\emptyset, \phi) \) closes iff \( \phi \) is valid (Theorem 3, p. 60). Translating the proof amounts to show that E will either manage, by systematically exploiting \( \Gamma \), to obtain the literal(s) requested in one history, or lead A to contradict himself, or will keep on playing if she cannot. In the first case, A will move ‘up’ in the tree until he can move ‘down’ again, and E will let him do so in a way that sequentially mimic systematic completed stT.

9 Notice however that attempted to use nonanalytic queries to reply A’s demands for literals is in general dominated by the strategic option of using analytic means (or lemmas). If A answers with the literal demanded, he facilitates E’s compliance with S.P. 2; if he answers with its negation, he does not, and oblige her to use analytic means, only now to force him into a contradiction. By S.P. 1, she should act as if expecting the latter, and therefore, that the nonanalytic strategy will merely delay the completion of the history by one exchange. By Ass. 1, she should in general favor the analytic one. However, obtaining a contradiction through a literal can sometimes minimize the number of future runs (because A will be aware of the consequences of this move, and won’t consider it as an option), hence may still be used in a way analogous to lemmas.
no guaranty that the game will be completed in some reasonable time (where ‘reasonable’
is a free parameter) when finite, which makes $O(s^*_a, s^*_e)$ a very inefficient strategy profile.

There is no mechanical procedure to find the ‘best’ lemma, unless entailment is already known to hold, hence the proof-theoretic importance of cut elimination. Lack of sophistication translates in lack of efficiency. Greater sophistication depends on procedures which address the problem of finding the ‘best’ lemmas, although they cannot solve it in general. Redundancies built in strategies in $O(s^*_a, s^*_e)$ suit players so unsophisticated that they cannot assign any resource to some such procedures. Humans computers are more sophisticated, and use cues—looking a variable-sharing, surface logical form, etc.—to obtain shortcuts.

4.2 EEG and GTS

A ‘bottom-up’ variant of GTS games can be obtained as a special case of EEG for some formula $\phi$ of $\mathcal{L}$, when $\Gamma = \emptyset$. First, one add an initial move from Nature, who chooses a model $\mathfrak{M}$. One also assumes that ranges of attention at a position, for both players, include any $(P\bar{a} \lor \neg P\bar{a})$ s.t. $(\neg)P\bar{a}$ has occurred at an earlier position in $h$ (in either a query or a defense). Informally, GTS is recovered in the special case when every atomic nonanalytic query is in player’s range of attention. The strategy profile that solves the game is as follows:

A plays an analytic strategy until either a literal is demanded from E. Then, E sends a nonanalytic query to Nature. E wins a run whenever Nature answers with the literal stated by E at the end of that run, and she has a w.s. if she can do so in every history.

This is tantamount to capture GTS games as a special case of Hintikka’s interrogative–deductive games (see Hintikka et al., 1999). Games of ‘pure discovery’ can also be captured, when the initial assumption that $\Gamma$ holds in the state of nature is correct, and when every answer is indeed true in the underlying state of Nature. Although in Hintikka’s model, A and Nature are identical, the discrepancy is is easily remedied by interpreting A as playing Nature’s role ‘by proxy,’ i.e. choosing answers whenever Nature does not provide one. Hence, a conclusion will be established as the outcome of the game if it holds in every situation that is indiscernible, from the players’ standpoint, from the state of Nature, given the information they have about it.

4.3 Learning entailment, and other relations

Although EEG do not decide first-order entailment, players can actually learn it, in the sense of formal learning theory (see e.g. Kelly, 2004). Even as unsophisticated player types those who find no better strategy than $s^*_a$ and $s^*_e$, can learn efficiently whether $\Gamma$ entails $\phi$. 
Consider e.g. a learning method that: (i) conjectures at $h_0$ that $\Gamma \not\models \phi$; and: (ii) changes its assessment iff the game is lost by $A$. Such a method clearly learns whether $\Gamma \models \phi$ or not with at most one retraction. Minimizing retractions is of particular importance since stabilization in the limit is compatible with erratic local behavior. More sophisticated players will learn quickly, but not more efficiently.\(^{10}\)

Modeling other entailment relation may require restrictions on $E$’s strategies (see e.g. n. 12). But it is also possible to model consequence relations where $A$ selects only within preferred subset of $\text{Mod}(\Gamma)$. The restriction on $A$’s strategy correspond in this case to strict preferences for choices of (at least some) choices of disjuncts or existential instantiations. These relations are typically non-monotonic, when the preferences express ‘default’ values, that can later be revised, allowing for representation of non-monotonic reasoning, provided that some device is added to represent this revision.\(^{11}\) If new values become available, new branches (runs) may become necessary, and whether such relations can be learned, depend on the one hand, on the stability of $\Sigma.P.$ in runs, and of the overall game tree.

\section*{5 Conclusion}

Our account of entailment games maps standard $\text{ST}$ building and closure rules, to (resp.) to local and global strategies of algorithmic players, playing under qualitative uncertainty, and with limited strategic insights. One systematic construction for $\text{ST}$ (complete for first-order entailment) is also characterized as the output of a strategy profile possibly selected by unsophisticated players. Unlike previous attempts in semantics inspired by game-theoretic notions, like $\text{DS}$ or extensions of $\text{GTS}$, the correspondence does not rely on ad hoc restrictions on players’ strategies. Moreover, our game-theoretic approach is versatile enough to capture (through adjustments in semantic clauses and the preference relations) other consequence relations than classical. It also generalizes and unifies the semantic models proposed by Hintikka for semantic and inquiry games, and captures formally algorithmic of what Hintikka’s informal ‘strategic principles.’

The main difference with Hintikka’s models is conspicuous when one tries to obtain from interrogative games a game-theoretic model of communication. In such a model, (i) Inquirer is a hearer (in a given context) attempting to interpret a speaker’s utterances; (ii)\(^{10}\) Notice that the above learning method is guaranteed to stabilize on the correct hypothesis provided that it is finite, or is carried indefinitely (in case of countermodels with infinite domains): it may output an incorrect estimate if ‘put on hold’ indefinitely, as may some inquiry games.

\(^{11}\) Hintikka et al. (1999) introduces ‘brackets’ to that effect, but does not discuss their strategic use. Genot (2009) presents a strategy that satisfies some of the Alchourrón-Gardenf axioms for contraction.
\( \Gamma \) is a ‘stock’ of speaker’s already interpreted past utterances; and: (iii) \( \phi \) is the postulated meaning of the current utterance. Interrogative moves provide feedback from speaker, rather than Nature, i.e. from a strategic player. In the absence of such feedback, Inquirer may be thought of as playing both A’s and E’s roles, relying on ‘preferred’ interpretations of (elements of) \( \Gamma \); unless the model makes unrealistic idealizations, Inquirer should not be construed as ‘cycling through’ possible interpretations, but rather stick with one—just as she currently test one candidate meaning \( \phi \).

A second difference is that feedback from speaker (answers) may be cooperative or not, depending on e.g. the motives of agents in the context of communication. On the one hand, in cooperative contexts, replies may go past the requested answers, if the speaker anticipates some further demands for information, or attempts to prevent misunderstandings, e.g. by explicit disambiguations. This may also trigger adjustments in \( \Gamma \), operating substitutions (revisions) or additions (expansions). On the other hand, in competitive argumentation, speaker’s position may be reinforced by maintaining ambiguity. In that case, hearer will be forced to consider more possible scenarios compatible with \( \Gamma \), but also multiple candidate interpretations \( \phi_1, \ldots, \phi_n \). Also, she must be ready to revise \( \Gamma \), but sometimes with the only option to remove what was taken for granted (contractions).

Clearly, the stronger the background assumptions, the least number of scenarios are to be considered. Given that, the EEG model is compatible with two interpretations. The first views logical reasoning as the basis of language interpretation, where stronger assumptions or habits lead to inductive reasoning by assuming ‘default’ meanings—assumptions which can be revised, making reasoning nonmonotonic. The second interpretation views logical reasoning as a cautious extension of linguistic interpretative practices. The first interpretation may appeal to logicians and rationalist philosophers, while the second is closer to evolutionary understanding of language.

We lean towards the latter interpretation, following e.g. Brinck and Gärdenfors (2003), who have argued that cognitive needs of cooperative communication have been a driving force behind the evolution of complex semantic representations. Logical reasoning through EEG is made possible by rather sophisticated manipulations of semantic representations: fixpoints are defined w.r.t. to regular extensions of semantic structures. We consider our model to be an abstract expression of the thesis that logical reasoning supervenes on language understanding; and we view generalizations of our games to nonclassical (and in particular, nonmonotonic) inference relations, as a step towards a semantic model of real-life interpretative practices.
Acknowledgments

Emmanuel J. Genot’s research was supported by Swedish Research Council postdoctoral research fellowship; Justine Jacot is a Ph.D. candidate at the University of Lund. Former versions of this paper were presented in March 2011 at the European Epistemology Network Meeting (Lund University); in June 2011 at the Center for Formal Epistemology, (Carnegie-Mellon University, Pittsburgh); and in June 2012, at the 6th workshop on Decision, Games and Logic (Center for Mathematical Philosophy, Munich). We thank participants for their comments and suggestions, in particular Peter Gärdenfors, Patricia Rich, Marta Sznajder and Ole Thomassen Hjortland. Finally, Horacio Arló-Costa had a deep and lasting influence on this work: it was him who suggested to make explicit the epistemic reasoning of players, and the role of uncertainty in it. His openness to others’ ideas, and the insights and erudition he willingly shared when discussing them, will be dearly remembered and sorely missed.

References


A Proofs

For Obs. 2, we leave implicit clauses other than for disjunction and existential quantifiers (and their negations)—and neglect double negation which has no optional argument under a classical reading.

Proof of Observation 2. Disjunction. A-case: By S.P. 2, A should avoid conceding many sentences E my obtain literals from as often as possible. By Ass. 2, one history may suffice for him. For those to reasons (in the later case, together with Ass. 1), A should not reply with both disjuncts. If the first run explored is not a win, A can try the other later (by Ass. 2 again). In the absence of particular insight, there is no reason for A to favor one disjunct over the other. E-case: When lacking insights about which option would best match A’s moves, E can assert both disjuncts, possibly ‘opting out’ one of them later, when information (given by A’s choices) increases. Clause (A2) of Def. 1 guarantees that E can retrace the past history, while Ass. 3 guarantees that she can retract one past choice.

Negated disjunction. A symmetric argument as for disjunction: A should attempt to minimize length of runs, and demand only one disjunct (possibly asking the other later); E should try to obtain as much explicit information as possible, and ask for both.

Existential quantifier. A-case: If L has infinitely many individual names, options for existential instantiation are nondenumerable (see Fig. 1). By S.P. 2, A should prefer introduction of new names (new individuals) in order to make harder E’s task to match literals he states. To avoid conflict with Ass. 1, A can introduce individual terms one at a time—AssAss:MeaningLog guarantees the possibility to introduce another later to expand his reply, if needed. E-case: By S.P. 2 E should prefer using ‘old’ names, previously introduced by A, and by Ass. 1, no more than necessary (which also makes ‘matching’ easier). Ass. 3 guarantees that she can rephrase or expand her reply at later stages, to match A’s choices. Hence, she should also reply with one instantiation at a time.12

Negated existential. A-case: One ‘new’ individual may suffice to prevent E from matching A’s choices (by S.P. 2), and avoids delaying runs unnecessarily (complying with Ass. 1; moreover, Ass. 3 guarantees A later ask for additional instantiations—in particular, if the only countermodels have infinite domains, A is

12 Multiple instantiations may be preferred when A or E has some particular insight, but a unique one is their best option when uncertainty precludes anticipations. E can introduce an individual name arbitrarily chosen, and later rephrase. Otherwise, she needs a ‘witness individual.’ Prohibiting ‘rephrasing’ replies for disjunctions and existential instantiation is one way to obtain intuitionistic logic (see Rahman and Keiff, 2005). This restriction can be derived from appropriate preferences (see also sec. 5).
still able to force an infinite play. E-case: By S.P. 2, E should prefer using ‘old’ names; and just as A in the existential case, by Ass. 1, should prefer to introduce them one at a time.\footnote{Again, E may realize, merely considering the past history, the need for multiple instantiations; still, they but they will be bounded by the number of names introduced by A. She can use queries over negated existential to force A to reply with names she has introduced, but such strategies necessary only if E does not choose her best option for existential instantiation. If rules are modified to allow for a single query to target n nested quantifiers, the number of options admissible for E will be the cardinal of the n-th Cartesian product of the set of already introduced individuals, and therefore will remain finite.}

For the proof of Obs. 4, we establish three lemmas, that show $\Sigma.P.$ to be stable over the three types of queries A can address without repeating one past move, i.e.: (a) rephrase a past reply with a new argument; (b) repeat a former (analytic) query with a new argument; or: (c) use a nonanalytic query.\footnote{Repetitions of queries with identical arguments are ruled out by Ass. 1 and S.P. 1: A should expect E to repeat her strategy, and such moves will delay the game unnecessarily.} In the first two cases, A can rephrase earlier replies or queries, w.r.t. to disjunctions and existential statements; changes in disjunct generate distinct histories, so we need only establish that $\Sigma.P.$ are stable in a given history over changes w.r.t. existential statement (negated or not).

**Lemma A.1** (Stability I). If, for some history $h$ of $\mathcal{G}(\Gamma, \phi)$ such that $P(h) = A$ the last position $m$ of $h$ is a $\Sigma.P.$, and if A retracts some past moves so that S.P. 2 is no longer satisfied at $m$, then E can modify her past moves as well, so that $m$ will be a $\Sigma.P.$ again.

**Proof of Lemma A.1.** If A can affect E’s compliance with S.P. 2 at $m$ by changing one of his past moves, then: (a) there is be some literal $(-)P\bar{a}$ (where $k_i$ occurs in $\bar{a}$) s.t. $E(-)P\bar{a}$ occurs at $(h|m_1)$ for $m_1 \leq m$, and $A(-)P\bar{a}$ occurs $(h|m_2)$ for $m_2 < m_1$; and: (b) A has substituted $A(-)P\bar{a}'$ to $A(-)P\bar{a}$ at $m_2$, with $\bar{a}' = \bar{a}[k_i/k_j]$, where $k_j$ is new. Ex hypothesis, A has played according to his best strategy when introducing $k_i$, hence $k_i$ was new at $m_2$. Moreover, $E(-)P\bar{a}$ cannot result from A’s imposing the use of $k_i$ (otherwise, he would have used $k_i$ twice, contrary to assumption that he conforms to preferences of Obs. 2). Hence, E needs simply to rephrase her instantiation with $k_i$ with $k_j$, and she will comply with S.P. 2 again. Then, both conditions of Def. 3 are met at $m$, which is a $\Sigma.P.$, as desired.\footnote{When $m_2 = m_3$, $\exists x\phi(x) = \chi = \exists P\bar{a}'$, where $\bar{a}' = \bar{a}[k_i/k_j]$.}

**Lemma A.2** (Stability II). If, for some history $h$ of $\mathcal{G}(\Gamma, \phi)$ such that $P(h) = A$, the last position $m$ of $h$ is a $\Sigma.P.$, and if A repeats some past query with a new argument, so that $h$ is extended into $h_1$, and some position $n_1$ of $h_1$ is no longer a $\Sigma.P.$, then E has a local strategy to extend $h_1$ to $h_2$ so that some position $n_2$ of $h_2$ is, again, a $\Sigma.P.$.

**Proof of Lemma A.2.** If A can affect E’s compliance with S.P. 2 at $m$, by reiterating a query with a different argument, then: (a) there are some positions $m_1$, $m_2$ and $m_3$, with $m_1 < m_2 < m_3 < m_4 < m_5 \leq m$ s.t. $E(-)\exists x\psi(x)$ occurs at $m_1$; $A(-)\exists x\psi(x)$ occurs at $m_2$; $A(\chi, (-)P\bar{a})$ (where $\chi$ is a subformula of $\psi[x/k_i]$, i.e. $k_i$ occurs in $\bar{a}$) occurs at $m_3$;\footnote{When $m_2 = m_3$, $\exists x\phi(x) = \chi = \exists P\bar{a}'$, where $\bar{a}' = \bar{a}[k_i/k_j]$.} $A(-)P\bar{a}$ occurs at $m_4$; and finally $E(-)P\bar{a}$ occurs at $m_5$. A’s reiteration consists in targeting $m_1$ with a new query $A(-)\exists x\psi(x)$ (where, according to his preferences, $k_j$ is new); then repeating the sequence until position $n_1$ is reached, where he repeats the query made at $m_3$, but this time with $(-)P\bar{a}'$, where $\bar{a}' = \bar{a}[k_i/k_j]$. At this position, the condition of Def. 3.b is no longer satisfied, and $n_1$ is no longer a $\Sigma.P.$ Let us call $h_1$ the history extending $h$ with this sequence of moves, with $n_1$ being the last position of $h_1$. $A(-)P\bar{a}$ (at $m_4$) must be the result of some sequence of moves from E, including at least one query $E(-)\exists x\theta(x)$ (where if not, $k_i$ is used twice by A, contrary to assumption that...}
he plays according to preferences); and also one query \(E(\sigma, (-)P\bar{a})\), where \(\sigma\) is a subformula of \(\theta[x/k_i]\), for if not, she cannot obtain \(A(\cdot)\overline{P\bar{a}}\), since \(A\) would not assert it as the result of one of his choices (if he does, he does not comply with S.P. 2, contrary to assumption). All \(E\) has to do is to extend \(h_1\) by appending to it a sequence of moves, starting with \(E(\neg\exists x \theta(x), x/k_j)\), until she can address the query \(E(\sigma', (-)P\bar{a}')\) (where \(\sigma' = \sigma[k_i/k_j]\)). \(A\) will eventually reply with \(A(\cdot)\overline{P\bar{a}'}\) (by assumption, \(A\) follows his best strategy; hence, replying at \(m_4\) was his best option at the time; and therefore, since the situation is identical save for the introduction of a new individual, replying must be his best option as well). And finally, \(E\) can reply to \(A\)’s query at \(n_1\) with \(E(\cdot)\overline{P\bar{a}'}\), in compliance with S.P. 2.

Let \(n_2\) denote the last position of \(h_2\). Clearly, both conditions of Def. 3 hold at \(n_2\), hence \(n_2\) is also a \(\Sigma.P\), as desired.

The last lemma covers the use of nonanalytic queries after a \(\Sigma.P\) has been reached. It follows from our assumptions, that \(A\) should be ready to answer a nonanalytic query that is ‘turned back’ against him, after finitely many steps; and is always indifferent between doing it immediately, or later.\(^{16}\)

**Lemma A.3** (Stability III). *If, for some history \(h\) of \(\mathcal{G}(\Gamma, \phi)\) such that \(P(h) = A\), the last position \(m\) of \(h\) is a \(\Sigma.P\), and if \(A\) extends \(h\) with some nonanalytic query and eventually to \(h_1\) where some position \(n_1\) is not a \(\Sigma.P\), then \(E\) can extend \(h_1\) to \(h_2\) so that some position \(n_2\) of \(h_2\) is, again, a \(\Sigma.P\).*

For simplicity the proof assumes that \(A\) answers immediately a nonanalytic query ‘turned back’ at him; extension to cases where he uses several in a row before he answers is left to the reader.

**Proof of Lemma A.3.** Assume that \(A\) extends \(h\) into \(h_1 = (h, A?\psi \lor \neg \psi)\). The proof is then by cases. **Case 1:** \(A(\cdot)\psi\) occurs at some position of \(h\). \(E\)’s best response, given S.P. 2, is to extend \(h_1\) into \(h_{1,1} = (h_1, E(\cdot)\psi)\) (because it guarantees her choice to match \(A\)’s). If \(\psi\) is a literal, the last position of \(h_{1,1}\) is, again, a \(\Sigma.P\).

If it is not, given that the last position of \(h\) is a \(\Sigma.P\), and that repetitions would only induce a new \(\Sigma.P\). (by Lem. A.1 and A.2), \(A\)’s best response is to use \(A(\cdot)\psi, \psi\) (where \(\psi\) is a nonempty argument when \(A\) can constrain the reply). But again, \(E\)’s best response is \(E(\cdot)\psi, \psi\), then to wait for \(A\)’s reply. Given that the ‘copycat’ strategy is \(E\)’s best response, the exchange is bound to proceed until a literal subformula \(\cdot)\overline{P\bar{a}}\) of \(\cdot)\overline{\psi}\) is reached—which, by \(E\)’s strategy, is then ‘copied’ from \(A\).\(^{17}\)**

**Case 2:** \(A(\cdot)\psi\) does not occur at some position of \(h\). \(E\) can extend \(h_1\) into \(h_{1,2} = (h_1, E?\psi \lor \neg \psi)\). By the same reasoning as in Case 1,

\(^{16}\) By Ass. 1, \(A\) should use \(A?\psi \lor \neg \psi\) at the last position of \(h\) (when \(P(h) = A\)) when it offers a prospect of reaching a \(\Sigma.P\). faster (given \(A\)’s insight at \(h\)). Therefore, if \(E\) plays \(E?\psi \lor \neg \psi\), \(A\) will typically not use all moves available prior to \(A?\psi \lor \neg \psi\) to delay a reply to \(E?\psi \lor \neg \psi\). Even if he does, the number of such moves without repetition is finite, and repetitions are useless to reach a \(\Sigma.P\). faster: they either delay it, or induce no change, by Lem. A.1 and A.2, and in both cases are ruled out by which is Ass. 1. Once the relevant moves are all spent, \(A\) can delay the reply to \(E?\psi \lor \neg \psi\) by using finitely many nonanalytic queries \(A?\psi \lor \neg \psi_2\), . . . , \(A?\psi \lor \neg \psi_n\). Doing so indefinitely would contradict S.P. 3 for sure, while a finite strategy may still be available; and every finite strategy requires him to be ready to answer any \(E?\psi \lor \neg \psi_i\) if asked to (by Ass. 3). Moreover, the prospect is no better to answer right away to \(E?\psi \lor \neg \psi_i\), or to line up \(A?\psi \lor \neg \psi_2\), . . . , \(A?\psi \lor \neg \psi_n\) before (eventually) replying to any one \(E?\psi \lor \neg \psi_i\).

\(^{17}\) If \(A\) has played more than one nonanalytic query, \(E\) can iterate the copycat strategy, until a \(\Sigma.P\) is reached. Notice also that \(E\) departs from preferences of Fig. 2 (replying with both disjunct, and retracting one later) because uncertainty about \(A\) future choices is not relevant.
A will eventually reply, and E's best response will then be to copy his reply. If ψ is a literal, the resulting position will be a Σ.P. (unless A has played more several queries, in which case E can iterate the copycat strategy, until a Σ.P. is reached); if not, as in Case 1, E has a (recursive) copycat strategy to reach a literal, and copy it. Let h₂ be the history whose last position is E(¬)Pₐ, and let n₂ denote that position. E(¬)Pₐ is copied on A(¬)Pₐ occurring before n₂, which and complies with S.P. 2, which satisfied Def. 3b. Moreover, every new E-labeled statements occurring between the last position m of h and n₂ has been targeted once, and ex hypothesis Def. 3a were satisfied at m, it is also at n₂. Hence, n₂ is a Σ.P., as desired. 

□

Obs. 4 follows from Lem. A.1–A.3, as follows:

Proof of Observation 4. Let |Pₐ|ₐₙ = {Pₐ : A(¬)Pₐ occurs in h} and |Pₐ|ₐₙ = {Pₐ : E(¬)Pₐ occurs in h}; let also m be the last position of h, and assume that it is a Σ.P.. Ex hypothesis (by Def. 3b) |Pₐ|ₐₙ ⊆ |Pₐ|ₐₙ. Either |Pₐ|ₐₙ is a complete diagram for some model of Γ, and identical with |Pₐ|ₐₙ, or not. In the first case, there is nothing to prove. In the second, A has to compel E to commit to a complete diagram |Pₐ|ₐₙ for some extension h′ of h. But A can obtain some (¬)Pₐ ∈ |Pₐ|ₐₙ only by repetition of some (sequence of) analytic query (queries) with the appropriate individual names; or by a nonanalytic query A?Pₐ ∨ ¬Pₐ.

It follows from (resp.) the proofs of Lem. A.2 and Lem. A.3, that in either case E has a recursive strategy that guarantees that she will obtain (¬)Pₐ from A, and therefore that, for any h′ t.s. |Pₐ|ₐₙ is a complete diagram, |Pₐ|ₐₙ = |Pₐ|ₐₙ, as desired. □

Proofs of Obs. 5 and 6 are direct consequences of Obs. 4 and S.P. 1–3.

Proof of Observation 5. Assume that the last position m of h is a Σ.P.. By Obs. 4, there is a recursive (partial) strategy sₗ t.s., in any extension h′ of h that is generated by sₗ (whatever A’s strategy sₗ is) where |Pₐ|ₐₙ is a diagram, then |Pₐ|ₐₙ = |Pₐ|ₐₙ. Build definition of (zₗ), h′ ∈ h″ h′ is an extension of h where E plays some other strategy s’ₗ. By S.P. 1, A should act as if expecting E will implement sₗ from m on. A’s only remaining strategy is to postpone indefinitely a comparison. Unless losing incurs for A an infinite loss, S.P. 3 applies, and A should prefer finite strategies to infinite ones. Since he looses in all the finite strategies that extend h, he should prefer the shortest (by Ass. 1), so that A’s best option is to extend h to h′ = (h, A-stop), as desired. □

Proof of Observation 6. We prove the case where Pₐ = Pkᵢ, and leave the others as an exercise to the reader. Assume (without loss of generality) that A Pkᵢ has occurred in h; and that the last position of h is E(ψ, ¬Pkᵢ). E has obtained A Pkᵢ, through a sequence of moves, as a subformula of either some A 3xψ’(x), in which case kᵢ was ‘new’; or of A¬3xψ’(x), in which case kᵢ was not. In both cases, the only way for A to retract Pkᵢ, is to substitute the first occurrence of kᵢ with some kⱼ. But, by the same reasoning as in the proof of Lem. A.1, E can restate her queries with kⱼ as well. Moreover, by S.P. 1, A should expect E to play the most harmful strategy for him, namely to repeat E(ψ, ¬Pkᵢ), until he answers. Unless the prospect of loosing is infinite, S.P. 3 applies, and A should prefer a finite strategy. Since in every finite strategy, he either plays eventually A¬Pₐ or A-stop, they all have the same prospect, and by Ass. 1, he should prefer the shortest. Hence, A’s best strategy is to extend h to h′ = (h, A-stop), as desired. □