



# Logical Dialogues with Explicit Preference Profiles and Strategy Selection

Emmanuel J. Genot<sup>1</sup>  · Justine Jacot<sup>1</sup>

Published online: 20 May 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** The *Barth–Krabbe–Hintikka–Hintikka Problem*, independently raised by Barth and Krabbe (From axiom to dialogue: a philosophical study of logics and argumentation. Walter de Gruyter, Berlin, 1982) and Hintikka and Hintikka (The sign of three: Peirce, Dupin, Holmes. In: Eco U, Sebeok TA (eds) Sherlock Holmes confronts modern logic: Toward a theory of information-seeking through questioning. Indiana University Press, Bloomington, 1983), is the problem of characterizing the strategic reasoning of the players of dialogical logic and game-theoretic semantics games from rational preferences rather than rules. We solve the problem by providing a set of preferences for players with bounded rationality and specifying strategic inferences from those preferences, for a variant of logical dialogues. This solution is generalized to both game-theoretic semantics and orthodox dialogical logic (classical and intuitionistic).

**Keywords** Game-theoretic semantics · Dialogical logic

## 1 Introduction: An Open Problem

In the 1930s, Wittgenstein articulated the view that logic is one of the many language games that can be played in natural or formal languages. He also suggested an analogy between having a proof and winning a game. *Dialogical logic* and *game-theoretic semantics* have provided formal interpretations of this analogy. Dialogical logic defines 2-player games in which *Proponent* has a winning strategy for defending a formula against the attacks of *Opponent* if the formula follows deductively from a (possibly

---

✉ Emmanuel J. Genot  
emmanuel.genot@fil.lu.se

Justine Jacot  
justine.jacot@fil.lu.se

<sup>1</sup> Department of Philosophy, Lund University, Lund, Sweden

empty) set of premises. Game-theoretic semantics defines 2-player games in which *Eloise* (*Abelard*) has a winning strategy for a formula if the formula is true (false) in an underlying model. [For an overview of dialogical logic, see [Rahman and Keiff \(2005\)](#); for game-theoretic semantics, see [Hintikka and Sandu \(1997\)](#)]. In both cases, the games are defined by sets of rules that specify explicitly the action sets of the players, and implicitly restrict the admissible strategies to those that emulate systematic proof construction and model-checking procedures.

Specifying a game by means of sets of rules is a slight departure from the standard practice in game theory. ‘Game rules’ are indeed either reduced to definitions (action sets) or to the players’ best response to some game configuration (strategic inferences). A game is defined by a set of players, a set of actions available to these players, and a preference relation for each player over the outcomes of their actions, that is very often represented by a payoff (utility) function for each player from their action set onto the set of real numbers. Of particular importance here is the absence of “rules” properly speaking. The underlying assumption of players’ rationality suffices to explain how to choose actions throughout the game by strategic reasoning over preferences, and when to stop the game. However, it is easy enough to translate these legal actions and strategic reasoning into rules enforcing certain preferences and good inferential practice. For instance, in the game of chess, rules for moving pieces define legal actions, while the checkmate rule expresses when it is rational to quit moving the king around the board because the payoff will not change.

Alternative formulations of dialogical logic and game-theoretic semantics based on preferences or utilities have been investigated in the 1980s respectively by [Barth and Krabbe \(1982\)](#) and [Hintikka and Hintikka \(1983\)](#). Barth and Krabbe suggested that rules that govern the end-game configuration of a dialogical game translate the preferences of rational arguers. They formulated closing rules in terms of what is rational for arguers to do, namely: the losing party has to acknowledge that the other party has won by rational means, and therefore acknowledges that it is irrational to keep arguing ([Barth and Krabbe 1982](#), p. 71). This explicitly introduces rational preferences for conceding defeat under particular circumstances in a play (as opposed to continuing arguing and extending that play). Rather than being told when to stop by rules, as in orthodox dialogical logic, Barth and Krabbe’s rational arguers stop arguing because they agree that one player has won. However, Barth and Krabbe did not characterize preferences over outcomes, and thus could not fully illuminate the relation between the preferences and the rules [see [Jacot et al. \(2016\)](#) for further details].

[Hintikka and Hintikka \(1983\)](#) hit closer to the mark, when they characterized the utility function for the construction of a tableau proof played relative to an underlying first-order model. The Hintikkas attached an incremental cost for each new individual name introduced in a proof that could not be assigned a denotation in the underlying model, yielding costs for both regular tableau construction and pure model checking as limit cases. The utility function formalized the intuition that sequences of moves that impose a lower load on working memory incur a lower cost, and expressed preferences over the outcome of a tableau-building process. Unfortunately, the Hintikkas formulated the utility function for *Player* in a game against *Nature*, and *Nature* is a non-strategic player whose moves are limited to providing information about its own state. Subsequently, the game is in fact a 1-player

game, that is, a sequential decision problem [see [Genot \(2017, Sect. 3.3\)](#) for further details]. The Hintikkas did not suggest how to extend their 1-player decision-making problem to a 2-player model-checking game *à la* game-theoretic semantics. Furthermore, maximizing expected utility is tantamount in this context to solving the Halting Problem ([Genot 2017, Sect. 2.3](#)), so that a decision-maker with a preference for low cognitive costs must perform hypercomputations in order to satisfy her preference.

Subsequently, Barth, Krabbe and the Hintikkas have left us with an open problem, namely specifying explicit preferences and strategy selection inferences for rational players of a dialogical logic game for validity, or a game-theoretic semantics game for truth in a model. Let us from now on refer to this problem as the *Barth–Krabbe–Hintikka–Hintikka* problem (BKHH for short). In order to solve this problem, one must characterize preferences over strategies, derived from preferences over the outcome of strategy profiles (pairs of strategies). The selection of a strategy by a player must depend on the desirability of the possible outcomes of that strategy for that player. And finally, the preferences and strategic inferences from those preferences must be suitable for reasoners with bounded computational resources.

In this paper, we solve the BKHH problem for a proprietary format of logical dialogues for classical first-order logic. This format is chosen so that our results apply to game-theoretic semantics and orthodox dialogical logic for classical and intuitionistic logic with minimal changes. Our format borrows heavily from dialogical logic, because the rules of the orthodox dialogical games can be interpreted as enforcing preferences for certain strategies. Our main contribution solves the difficulties mentioned above by characterizing those preferences over outcomes and the strategic inferences from those preferences that yield preferences over strategies.

The remainder of this paper is organized as follows. Section 2 presents signed semantic trees for logical games (2.1 and 2.2), develops our format of game trees for logical dialogues (2.3) and explains the differences with standard dialogues (2.4). Section 3 exposes a standard argument for the completeness of logical dialogues (3.1) and shows how to handle players with preferences (3.2). Section 4 turns to these preferences, first characterizing players' bounded rationality (4.1), and second articulating the form of strategic reasoning appropriate for them (4.2 and 4.3). Section 5 implements these preferences in logical dialogues, provides representation results for players' reasoning to optimal choices of moves (5.1), and for players' reasoning to end-of-play and end-of-game choices (5.2), before returning to completeness (5.3). Section 6 extends the solution to orthodox game-theoretic semantics (6.1), and orthodox dialogical logic for both the classical and intuitionistic cases (6.2). An appendix collects the proofs of the propositions discussed in Sects. 4 and 5.

## 2 Of Proof Trees and Game Trees

### 2.1 Preliminaries

Logical dialogues are 2-player games presented as two-column tableaux [one column for each player's moves, cf. [Rahman and Keiff \(2005\)](#) or [Keiff \(2009\)](#)], reminiscent of

semantic tableaux introduced by Beth (1955), which are also the formal representation of 1-player proof games discussed by Hintikka and Hintikka (1983). The completeness of logical dialogues is proved by mapping their extensive form (game tree) onto signed semantic trees, introduced in Smullyan (1968), which are in all effect single-column versions of Beth tableaux. Our main concern is with strategic reasoning in logical dialogues *qua* extensive games. Hence, we consider their tree form rather than their tableau form. The rest of this section assumes no familiarity with logical dialogues and a passing familiarity with semantic tableaux. Knowledgeable readers may skip it, with the exception of the last paragraph, which accounts for some deviations from standard semantic trees methods. As mentioned in the introduction, we reconstruct classical first-order logic, and will later generalize our solution.

Given a first-order language  $\mathcal{L}$  of arbitrary signature with its standard compositional semantics, a semantic tree formalizes an attempt at proving by *reductio* that some (possibly empty) set  $\Gamma$  of sentences of  $\mathcal{L}$  entails a single sentence  $\phi$  of  $\mathcal{L}$ , or that  $\phi$  is a logical truth when  $\Gamma = \emptyset$ . (Everything we say hereafter about semantic trees holds *mutatis mutandis* for two-columns Beth tableaux.) Tree-building rules: (1) interpret all elements of  $\Gamma$  as true, and  $\phi$  as false; and: (2) allow for decomposing elements of  $\Gamma$  and  $\phi$  into subformulas using rewriting rules that preserve this interpretation. A semantic tree is *analytic* if every formula that occurs at a position is a subformula of  $\phi$  or some element of  $\Gamma$ , in which case the truth-value of any formula at any position depends uniquely on the truth-values assigned to elements of  $\Gamma$  and to  $\phi$ . Selective violations of the subformula principle are possible via the *Cut* rule (cf. Sect. 2.4). If  $\Gamma$  does not entail  $\phi$ , the rules yield at least one branch in which the assignment satisfies the *reductio* assumption (the branch is said to be *open*); and if  $\Gamma$  entails  $\phi$ , the rules generate contradictory assignments in every branch (the branches are said to be *closed*, and so is the tree). The rules can be applied systematically so that every closed branch is finite, but some open branches may be infinite (see Smullyan 1968, p. 59).

Formally,  $\Gamma$  and  $\phi$  are placed at the root of a proof tree, and tree-building rules are then applied to a formula at a given node  $n$ , generating new nodes, possibly in parallel branches.  $\Gamma$  is listed when finite, otherwise, the systematic methods rely on enumerations of the premises that guarantee that none of them will be omitted. Semantic trees come in two flavors, *signed*, and *unsigned*. In *signed semantic trees*, any formula that occurs at a node of the tree is ‘signed’ with one truth-value, *true* (**T**) or *false* (**F**). Specifically, every  $\gamma \in \Gamma$  introduced in the proof tree is **T**-signed, whereas (by the *reductio* hypothesis),  $\phi$  is **F**-signed, and tree-building rules reflect this initial assignment. *Unsigned* semantic trees dispense with explicit truth values, by placing the negation  $\neg\phi$  of the conclusion  $\phi$  at the root, while elements of  $\Gamma$  are introduced as they are. The tree-building rules incorporate duality laws, i.e. push negation in. In either type of trees, the number of successor nodes to a node  $n$ , generated by a tree-building rule applied to that node, is always finite. However, some rules can be applied iteratively, and thus some branches may have infinitely many nodes. In technical jargon, semantic trees are *finitely generated* but are not always finite.

The next section introduces a ‘hybrid’ of signed and unsigned semantic trees. The purpose of this modification is to make possible to solve the BKHH problem for both game-theoretic semantics and orthodox dialogical logic, which we will do in Sect. 6.

For now, suffice to say that this modification is intended to obtain logical dialogues where the commitments of the players are kept apart. The first assertion of a logical dialogue is the *thesis* asserted by *Proponent* (**P**), and mapped to the conclusion  $\phi$  of a proof tree.  $\Gamma$  (when not empty) is treated as a set of concessions of the *Opponent* (**O**) to the thesis that **P** can use to back her assertion. Therefore, the *thesis* is mapped to a **F**-signed formula, and the premises to **T**-signed formulas. Legal moves are specified as attacks and defenses, and may result in **P** (respectively **O**) attacking a **O**-labeled (**P**-labeled) formula  $\psi$  by asserting a subformula of  $\psi$ . This sequence would be mapped to tree-building rules that change sign between a formula at one node and some of its subformulas at successor nodes. However, assuming duality laws, legal moves for logical dialogues can also be specified so that **P**'s assertions depend only on  $\phi$ , and **O**'s assertions, only on  $\Gamma$ . This is tantamount to using both the **T**-signed and **F**-signed formulas *and* rewriting rules to 'push negation in'. We examine some consequences relative to logical dialogues in Sect. 2.4, and come clean about our motivations in Sect. 6.1.

### 2.2 Signed Semantic Trees (A Solitaire Game)

Tree-building rules for a language  $\mathcal{L}$  with propositional operators  $\neg$  (negation),  $\wedge$  (conjunction) and  $\vee$  (disjunction) are represented in Fig. 1 (where  $\psi_1$  and  $\psi_2$  are formulas of  $\mathcal{L}$  of arbitrary complexity). **T** and **F** 'sign' formulas (resp.) as *true* and *false* (as with *signed* semantic trees), and prefixed negation operators distribute over formulas in their scope according to duality laws (as with *unsigned* semantic trees). The material conditional can be introduced by the usual definition  $\psi_1 \rightarrow \psi_2 \stackrel{\text{def}}{=} \neg\psi_1 \vee \psi_2$ , and is omitted. Although the rules reflect the compositional semantics for  $\mathcal{L}$ , they are more economical than semantic clauses. For instance, for a true disjunction **T** $\psi_1 \vee \psi_2$ , there is no branch for the alternative where both  $\psi_1$  and  $\psi_2$  hold, and the same goes for its dual **F** $\psi_1 \wedge \psi_2$ . The rules mimic informal *reasoning by cases*, where the third branch of the alternative can be omitted when the two others have been explored. However, the existence of a third alternative will matter for logical dialogues. Unrepresented alternatives also exist in the case of rules for existential ( $\exists$ ) and universal ( $\forall$ ) quantifiers, represented in Fig. 2, which introduce individual parameters in the proof.<sup>1</sup> These rules can in principle be applied arbitrarily many times, and their systematic application is critical for proving the completeness of semantic tableaux [again, see Smullyan (1968, p. 59)].

A branch  $b$  generated by successive applications of tree-building rules of Figs. 1 and 2 is said to be *open* if the constraints on the assignment of truth-values to **T**- and **F**-signed formulas occurring in  $b$  are jointly satisfiable. Otherwise, the branch is said to be *closed*. More explicitly, if for some formula  $\psi$ , both **X** $\psi$  and **X** $\neg\psi$  or both **X** $\psi$  and **Y** $\psi$  occur in  $b$  (with **X**, **Y**  $\in$  {**T**, **F**}, **X**  $\neq$  **Y**), then the rules impose

<sup>1</sup> Parameters can be interpreted as their own domain, in which case there is no need to assume an underlying domain (*substitutional* semantics), or as variables of a special type that always occur free, and can be assigned denotations in an underlying domain of objects (*objectual* semantics). The former is more natural in dialogical logic (no underlying domain is presupposed), and the latter is more natural in game-theoretic semantics (an underlying domain is presupposed).

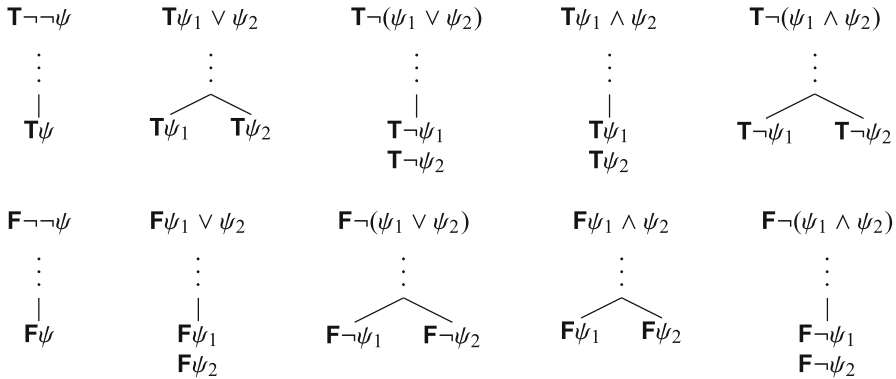


Fig. 1 Building rules for hybrid trees—operators

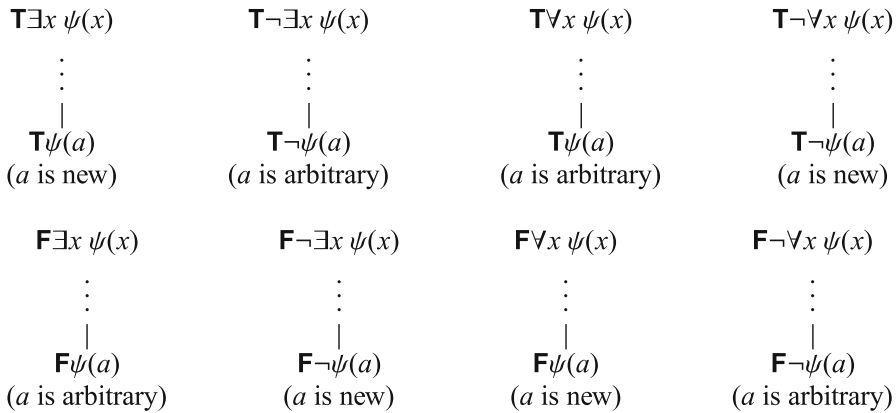


Fig. 2 Building rules for hybrid trees II—quantifiers

constraints that are not jointly satisfiable according to the semantics of  $\mathcal{L}$ . *Closure rules* then halt the construction of a branch because closed branches cannot yield counterexamples. When all branches are closed, there is no admissible assignment under the initial assumptions: the whole tree is declared *closed* by extension, and (by the *reductio* assumption)  $\Gamma$  entails  $\phi$ .

A theorem prover has some leeway for applying tree-building and closure rules relative to some particular  $\Gamma$  and  $\phi$ . Different proof strategies can result in proofs of different length and complexity, and there is no general solution to the problem of finding the best strategy to prove whether  $\Gamma$  entails  $\phi$  or not.<sup>2</sup> However, the problem of finding a strategy that outputs a closed tree whenever  $\Gamma$  entails  $\phi$  has a solution. This solution guarantees that a tableau with premises  $\Gamma$  and conclusion  $\phi$  closes iff  $\Gamma$

<sup>2</sup> The notion of ‘strategy’ in theorem-proving is defined by Smullyan (1968). Intuitively, the optimal strategy is the strategy that applies the rules to the smallest set of premises necessary to close the tableau. Assume now that there is a strategy for building a closed tableau with conclusion  $\phi$  in a tableau with premises  $\Gamma$ , and that there is a subset  $\Gamma' \subset \Gamma$  such that  $\Gamma'$  does not entail  $\phi$ . If the theorem prover reasons by elimination of dominated strategies, she must be able to eliminate every strategy where she applies tableaux

entails  $\phi$ . In technical jargon, semantic trees are *sound and complete* with respect to the semantics of first-order logic. Moreover, the solution is constructive: completeness proofs can exhibit systematic strategies that yield a finite closed tree with  $\Gamma$  and  $\phi$  at its root whenever  $\Gamma$  entails  $\phi$ . Insights into proof construction may yield more elegant proofs than those obtained by systematic constructions, but they are not necessary because these systematic constructions can be implemented mechanically. Building a proof tree can be compared to a game of solitaire, with the aim of learning whether or not some set of premises  $\Gamma$  entails  $\phi$ , and is a ‘language game’ insofar as the ‘game rules’ (the tree-building and closure rules) reflect the semantics of  $\mathcal{L}$ . Proof trees, therefore, offer a first approximation of a formal interpretation of Wittgenstein’s game metaphor (Jacot et al. 2016).

### 2.3 Logical Dialogues (A Two-Player Game)

Logical dialogues are akin to a *pro-and-contra* argumentation, with an immediate relation to proofs: *Proponent* (**P**) is committed to prove that thesis  $\phi$  follows from some  $\Gamma$ . Symmetrically, *Opponent* (**O**) is committed to prove that  $\phi$  does not follow from  $\Gamma$ . Whichever player fulfills successfully their initial commitment according to the rules wins the game. These rules are of two types, *particle rules* and *structural rules*. Particle rules encode the semantics of the language in which the game is played and contribute to determine legal moves—attacks and defenses—that follow **O**’s initial concession of  $\Gamma$ , and **P**’s initial statement of  $\phi$ . Following the Wittgensteinian *motto* that “meaning is use”, particle rules specify the meaning of logical operators according to how they can be used in the game, and govern what player **X** can ask player **Y** following **Y**’s assertion of some sentence  $\psi \in \mathcal{L}$ , based on the main operator of  $\psi$ . Equivalently, they express the constraints on **Y**’s future statements, imposed by **Y**’s statement of  $\psi$ . Structural rules govern the general set-up of the game, such as the order of play and the winning conditions, but also further restrict the legal attacks and defenses, and thus are best discussed after the particle rules have been introduced.<sup>3</sup>

Figure 3 presents rules for attacks and defenses with the following conventions: **X** $\psi$  represents a position where **X** has stated  $\psi$ ; **Y**? $\psi$  represent **Y**’s attack against  $\psi$ ; options for attacks and defenses determine ‘branching’ histories. If the rule allows **Y** to constrain the defense, a branching results from **Y**’s move and the constraint is specified between ‘<’ and ‘>’; otherwise, **X** retains the options for defenses. In either case, the options determine equally many branching histories. As previously mentioned, dialogical logic typically does not keep a strictly parallel track of **O**’s and

Footnote 2 continued

rules only to  $\Gamma'$ . Assume now that the only models of  $\Gamma'$  where  $\phi$  does not hold are infinite. Then all the open branches of a tableau with premises  $\Gamma'$  and conclusion  $\phi$  are also infinite. The strategic problem of eliminating dominated strategies where the theorem prover applies tableau rules only to  $\Gamma'$  is equivalent to the *Halting Problem*, which is not effectively solvable.

<sup>3</sup> **O** and **P** need not have any understanding of the rules for attacking and defending assertions, or the structural rules, other than the understanding needed to apply them correctly. In particular, truth-functional semantics is not assumed to be in the common ground when playing with the set of rules for classical logic, no more than any semantic interpretation of intuitionistic logic, when playing with the rules for that logic.

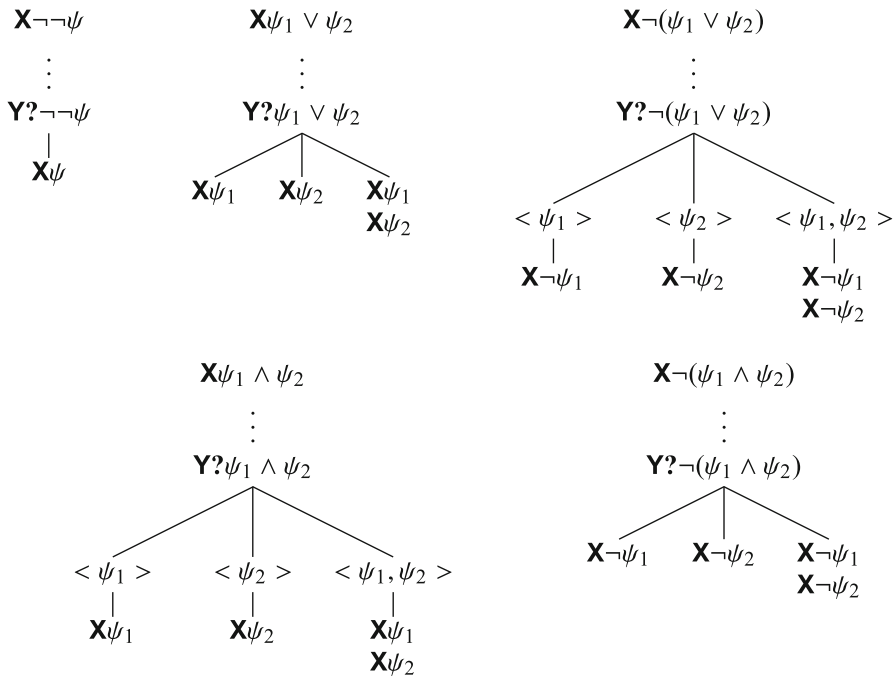


Fig. 3 Particle rules I: propositional connectives

**P**'s commitments: **Y** can attack  $X\neg\neg\psi$  with  $Y\psi$  and  $X\psi_1 \rightarrow \psi_2$  with  $Y\psi_1$ . We will discuss some consequences of the modifications we have introduced in Sect. 2.4, but a complete explanation of the reasons behind our choice will be delayed until Sect. 6.

Particle rules are not among the ‘rules’ that a solution to the BKH problem need to get rid of. The reason is that they are merely a representation of the players’ *action set* in the game. However, the terminology of ‘tree-building rules’ and ‘particle rules’ is nicely parallel and well entrenched, and it is of no consequence for our argument to keep referring to the action set of players relative to formulas as ‘particle rules’ (we will use the phrase ‘action set’ on occasion as a reminder). Particle rules for binary connectives generate a topology that is rather similar to that of semantic trees, with the exception of the additional options for attacks and defenses. In particular, the tree remains finitely generated. However, the quantifier rules introduce major changes in the topology, as soon as  $\mathcal{L}$  has (countably) infinitely many parameters, that is when the set of parameters in  $\mathcal{L}$  is  $K = \{k_i : i \in \mathbb{N}\}$ .<sup>4</sup> In that case, if  $X\forall x\psi(x)$  occurs at a position, then **Y** can in principle ask **X** to commit to arbitrarily many instantiations of  $\psi(x)$ , or equivalently, chose *any subset* of  $K$  for **X** to defend  $X\forall x\psi(x)$  with. And similarly, if  $X\exists x\psi(x)$  occurs at a position in the game, and is challenged by **Y**, then **X** can commit to as many instantiations of  $\psi(x)$  as **X** wishes. When the set of options for attacks or defenses is the powerset of  $K$ , not only do game trees cease to be finitely generated,

<sup>4</sup> The assumption, which is not expressible in  $\mathcal{L}$ , and thus cannot be made explicit in  $\Gamma$ , must be assumed to be part of the common ground in the same sense as the grammar of  $\mathcal{L}$ .



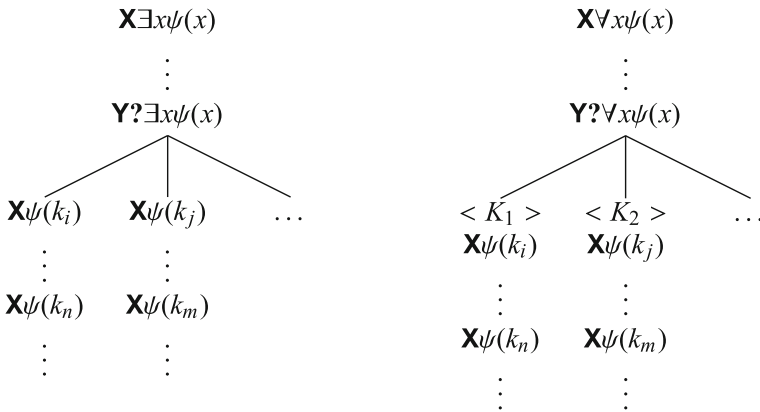


Fig. 4 Particle rules II: quantifiers

but the number of options becomes in principle uncountable. Figure 4 represents partially attacks and defenses for the unnegated existential quantifier, where **X** can choose  $K_i \subseteq K$ , and for the universal quantifier, where the choice of  $K_i$  is **Y**'s. The representation is partial because the set of options can be diagonalized (left to the reader). Rules for quantifiers within the scope of negation operators are omitted, but straightforward.

The number and order of structural rules tend to vary [Rahman and Keiff (2005) list six rules, and Keiff (2009), seven], but these differences are of little consequence for our exposition. Structural rules specify, *inter alia*, who plays first; under which condition a play is won; when and how **O** is authorized to start a new play (see Sect. 3); whether or not players are allowed to delay their defenses, to change them at a later stage, or to repeat some attacks. Thus, some of these rules govern the game set-up, some are merely definitions, and others actually restrict players' strategies. Our concern will be two rules of the latter kind. The first restricts **P**'s strategies by forbidding her to state an atomic formula of  $\mathcal{L}$  if **O** has not stated that atom at an earlier position, which Rahman and Keiff (2005, p. 369) list as fourth (SR-4 "formal use of prime formulae") and Keiff (2009) as fifth (SR-5 "Formal Use of Atomic Formulas"). Because the rules of Figs. 3 and 4 incorporate duality laws, our formulation is modified to cover both atoms and negated atoms, or *literals*, as such:

**Structural Rule–Use of Literals (SR-L):** **P** cannot state a literal sentence  $\psi$  of  $\mathcal{L}$  at a position of a play unless **O** has already stated  $\psi$  at an earlier position in that play.

The second structural rule of interest restricts 'strict' repetitions of attacks, which Rahman and Keiff (2005, p. 370) list as fifth (SR-5 "no delaying tactics rule") and Keiff (2009) as sixth (SR-6 "Classical No-Delaying-Tactics Rule"). The following informal formulation, which avoids the details of what counts as 'strict' repetition, will suffice for our purpose:

**Structural Rule–Repetitions (SR-R):** If player **X** has already attacked a statement  $\mathbf{Y}\psi$ , **X** cannot target  $\mathbf{Y}\psi$  again, unless: (1) the attack has an optional argument; (2) **X** chooses a new value for that argument; and: (3) the attack does not simply delay the application of another rule, in particular an end-of-play rule.

Other structural rules of interest are end-of-play rules, which enforce the closure of a play if: (1) **P** answers a literal attack from **O** and **O** cannot repeat the attack in compliance with (SR-R), and: (2) either **O** or **P** states a contradiction. Moves that open alternatives may give rise to alternative courses of the game, or *plays* (as referred to in SR-L). In keeping with simple intuitions about argumentation, a play goes to the player who has the last word in that play, but it does not always settle the game: winning conditions for **O** and **P** are asymmetrical, and winning a play is not sufficient for **P** but suffices for **O**. More precisely, a *winning strategy for P* is a strategy that responds to any sequence of attacks on the thesis  $\phi$ , using only the information that she can extract from elements of  $\Gamma$ . Equivalently, a winning strategy for **P** is a strategy that allows **P** to win *every play* that **O** can force her to play. Symmetrically, a *winning strategy for O* is one that generates (at least) one sequence of attacks such that **P** cannot answer them all without requiring more than what is conceded in  $\Gamma$ .

## 2.4 Remarks on Our Action Set for Logical Dialogues

We conclude this section with two remarks about the rules of Figs. 3 and 4. The first remark is that the main interest of those rules is to keep **O**'s and **P**'s commitments apart. Namely, with these rules, a statement is **O**-labeled (**P**-labeled) iff it is a subformula of some  $\gamma \in \Gamma$  (a subformula of  $\phi$ ). This property will allow us to extend our solution to the BKHH problem to orthodox game-theoretic semantics and dialogical logic. On the surface, our rules reflect a classical understanding of propositional connectives. And yet, while assuming duality laws and double negation, the rules of Fig. 3 are too weak to guarantee that **P** can win a dialogue where the thesis is a classical tautology. To see this, consider a logical dialogue with  $\Gamma = \emptyset$  and  $\phi = (p \vee \neg p)$  where  $p$  ( $\neg p$ ) is an atom (negative literal). By (SR-L), **P** cannot defend herself against an attack on  $\phi$ . In fact, the rules are also too weak for **P** to win a dialogue with  $\Gamma = \emptyset$  and either  $\phi' = (\neg\neg p \rightarrow p)$  or  $\phi'' = (p \rightarrow \neg\neg p)$ . The double negation rule is indeed not strong enough to win either  $\phi'$ , which is classically valid, or  $\phi''$ , which is intuitionistically valid.

There is an easy fix to this situation, which consists in allowing **P** to ask **O** for concessions of instances of the Excluded Middle. If **O** always answers, then **P** can win all classically valid tautologies.<sup>5</sup> If **O** can refuse to answer some of those questions, then

<sup>5</sup> *Sketch of proof:* assume that  $\phi \in \mathcal{L}$  is a classically valid tautology. Hence,  $\phi$  is interpreted as true in all assignments of truth values to its atomic subformulas. Conversely,  $\neg\phi$  is false in all those assignments. Hence, in a logical dialogue about  $\phi$  in which **P** is allowed to ask for instances of the excluded middle, **P** can ask for  $\phi \vee \neg\phi$ . If **O** answers  $\mathbf{O}\neg\phi$ , **P** can force him to state a contradiction. If **O** answers  $\mathbf{O}\phi$ , **P** can play a 'copycat' strategy, in which she attacks **O**'s statement that  $\phi$  with the same attacks, and in the same order as **O** has attacked **P** $\phi$ , and copies **O**'s defenses.  $\square$

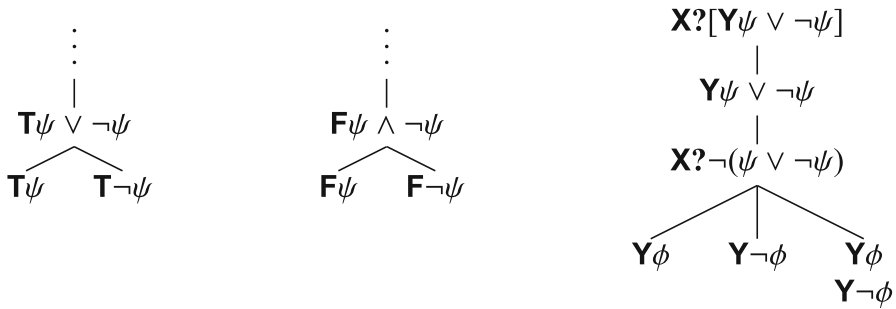


Fig. 5 Cut rule for tableaux and logical dialogues

**O** can deny victory over classical tautologies that are not constructively (intuitionistically) valid, although in our above examples,  $(\phi' \vee \neg\phi')$  should not be conceded but  $(\phi'' \vee \neg\phi'')$  should. However, we need not be concerned with this complication, since the introduction of instances of the Excluded Middle is necessary to extend our solution to game-theoretic semantics, while the extension to logical dialogues for intuitionistic logic will only assume the action set specified by standard particle rules. As for the addition to the players' action set, it is the counterpart of the *Cut Rule* for tableaux systems, represented in Fig. 5 on the left-hand side, with the corresponding sequence of moves for logical dialogues on the right-hand side, where  $X?[Y\psi \vee \neg\psi]$  represents the demand by **X** of a concession of  $(\psi \vee \neg\psi)$  from **Y**.

Our second and last remark concerns the departure of the particle rules of Fig. 4 from the quantifier rules given in standard expositions of logical dialogues such as [Rahman and Keiff \(2005\)](#); [Keiff \(2009\)](#). Standard particle rules restrict the legal strategies to those that introduce one parameter at a time, with (SR-R) allowing for multiple attacks under the condition that the optional parameter is new. Hence, standard particle rules *alone* do not actually specify the meaning of logical operators: in the case of quantifiers, meaning is actually given by the interplay of particle *and* structural rules.

### 3 Completeness of Logical Dialogues

#### 3.1 The Standard Completeness Argument

Logical dialogues are *complete* for classical first-order logic if for any first-order language  $\mathcal{L}$ ,  $\Gamma \subseteq \mathcal{L}$ , and  $\phi \in \mathcal{L}$ , the two conditions are equivalent:

1.  $\Gamma$  entails  $\phi$ , that is: all the models of  $\Gamma$  are models of  $\phi$ .
2. **P** has a *winning strategy* in a logical dialogue where the thesis is  $\phi$ , and where **O** concedes  $\Gamma$  as part of the common ground.

Alternatively, condition (1) can be formulated in terms of syntactic consequence, relative to the set of classical theorems. The above formulation is however more intuitive, because the equivalence of conditions (1) and (2) is established by mapping the game trees for logical dialogues to signed semantic proof trees. Intuitively, game trees

for logical dialogues where **P** plays her ‘best’ strategy against **O** are equivalent (*modulo* the mapping) to proof trees resulting from the implementation of some procedure that guarantees a closed tree exactly when  $\Gamma$  entails  $\phi$ . This section summarizes the procedure by which the mapping is obtained. Formal details can be found in [Rahman and Keiff \(2005, Sect. 2, pp. 371–375\)](#).

The first step of the mapping relies on an informal argument about **O**’s and **P**’s preferences between alternative moves, based on their respective goals. Schematically, **O** is better off selecting attacks that force **P** to obtain as many concessions as possible, and defenses that concede as little as possible. Symmetrically, **P** is better off asking as much as possible, and choosing defenses that could be backed with fewer concessions. Subsequently, **O**’s and **P**’s best options among the player-independent options of Figs. 3 and 4 correspond, respectively, to **T**-cases and **F**-cases for semantic trees of Figs. 1 and 2. Therefore, the local topology of a proof tree and the local topology of a logical dialogue game tree are equivalent to one-another *modulo* the substitution of **O** to **T** and of **P** to **F**, and the omission of ?-prefixed nodes.

The first step does not guarantee that a *play* of a logical dialogue terminates with a victory for **P** exactly when the corresponding *branch* in the semantic tree closes. The second step of the mapping is thus to establish a correspondence between closure rules and end-of-play rules, and faces two minor issues, both related to (SR-L). First, closure rules for signed semantic trees apply to pairs of sentences of arbitrary complexity, but (SR-L) prevents logical dialogues to stop until atoms (or in our formulation, literals) are reached. Second, (SR-L) prevents **P** to state  $\mathbf{P}\psi$  in a play for some atom (literal)  $\psi$  if  $\mathbf{O}\psi$  has not occurred in that play, but  $\mathbf{F}\psi$  can be obtained with tree-building rules in a branch even when  $\mathbf{T}\psi$  has not occurred in the same branch. The first issue is solved by appeal to the proof that every closed branch of a signed semantic tree can be extended to an *atomically* closed branch ([Smullyan 1968, p. 47](#)), and remarking that plays of logical dialogues won by **P** map to *atomically* closed branches. The second is solved by showing that omission of **T**-signed formulas corresponding to statements that (SR-L) prevents **P** to make is never sufficient to open a closed branch ([Rahman and Keiff 2005, Sect. 2](#)). Therefore, a branch-to-branch correspondence is established: the branch of a signed semantic tree is closed if the play it is mapped to terminates with **P**’s victory.

A third step is needed to extend the branch-to-branch correspondence to a tree-to-tree correspondence, and its importance is easily overlooked. Learning whether  $\Gamma$  entails  $\phi$  is always parasitic on the full tree, be it a proof tree or a game tree.<sup>6</sup> In the solitary version of the game, there is no difference between a play and a game, because systematic strategies build branches in parallel. But in a logical dialogue, the outcome of a play is a *branch*, and since learning whether  $\Gamma$  entails  $\phi$  supervenes on learning whether or not **P** has a winning strategy in the whole *game*, multiple plays have to

<sup>6</sup> Learning whether  $\Gamma$  entails  $\phi$  is possible even when the proof tree or game tree is infinite. In semantic tableaux, this is a consequence of the *reductio* assumption, since one conjectures that  $\Gamma$  does not entail  $\phi$ . If the tree construction never ends, one’s conjecture remains correct; if the tree construction ends and the tree is open, it is also correct; if it terminates and the tree is closed, it is not, but one can change one’s mind. Similarly, in a dialogical game, the initial conjecture is that  $\Gamma$  does not entail  $\phi$ , which is why **P** has the burden of the proof. For a formal explication of the appropriate notion of ‘learning’ in that context, see [Kelly \(2004\)](#).

be played. A technical solution is to define so-called “strategy games” (Rahman and Keiff 2005, pp. 371–375) that introduce an ad hoc rule, taking effect when a *play* terminates with a victory for **P**. The rule allows **O** to go back to the last position where **O** had to select an option for either an attack or a defense, and to explore any option left unexplored. This rule is actually equivalent to stipulating that **O** and **P** play a *sequence of plays*, keeping memory of the past plays, and play as many plays as necessary to assess whether **P** has a winning strategy or not. Together with the first two steps, this last step is sufficient to complete the mapping of game trees to signed semantic trees.

### 3.2 Some Difficulties (and How to Solve Them)

Game trees for “strategy games” are constructed sequentially, while the construction of signed semantic trees is parallel, which complicates further the mapping between the two.<sup>7</sup> The issue is however minor, as shown by the completeness proof proposed by Rahman and Keiff (2005), which maps the logical dialogues game trees to the output of a ‘sequential’ (depth-first) algorithm to build semantic proof trees, rather than the usual parallel (breadth-first) one. A more serious difference is that the systematic construction of a signed semantic tree is a mechanical task for a single agent, whereas the game tree of a logical dialogue represents the outcome of the interaction between two agents. In order to solve the BKHH problem, one must explain why a game tree for a logical dialogue where the players best respond to one another’s strategy maps to a signed semantic tree generated by a (depth-first) systematic method.

This task can be solved by ‘reverse engineering’ preferences from logical dialogues. First, dialogical logic explicitly defines preferences that affect the local topology of the game tree (why **O** and **P** prefer different attacks and defenses). These preferences are not yet sufficient to explain the global topology of the tree (why **O** and **P** would select a *pair of strategies* realizing a proof tree). But (SR-L) and (SR-R) impose constraints at the intermediate level of branches that are equivalent to rational preferences for arguers in the sense of Barth and Krabbe (1982). For instance, the strongest possible justification an arguer can have for a claim in a *pro-* and *contra-*argumentation, is a concession from her opponent with the same content. And (SR-L) is nothing if the formal equivalent for that justification. Barth and Krabbe also suggest that rational arguers should seek victory by ‘rational means’ alone and avoid delaying tactics. This excludes in particular delaying tactics to avoid loss, which (SR-R) prohibits. Hence, both (SR-L) and (SR-R) could as well be ‘self imposed’ by rational arguers. Subsequently, one can treat compliance with (SR-L) and (SR-R) as *revealed*

<sup>7</sup> Assume that  $\Gamma$  does *not* entail  $\phi$ , and that there is only one infinite countermodel to  $\phi$ . A systematic signed semantic tree with root  $\Gamma$  and  $\phi$  will have one infinite branch, but also finitely many closed branches corresponding to models of  $\Gamma$  satisfying  $\phi$ . Furthermore, if  $\Gamma$  is finite, the systematic procedure will generate all the closed branches after finitely many iterations, and if  $\Gamma$  is infinite, the procedure will generate them all in the limit. However, a logical “strategy game” dialogue will lock **O** and **P** in an infinite play as soon as **O** figures out how to introduce infinitely many parameters in agreement with (SR-R), which is in fact tantamount to building an infinite countermodel (see Sect. 6). If it is the first play of the game, then the sequential construction of the game tree will never converge to the parallel construction of the systematic signed semantic tree.

*preferences* for (local) strategies derived from preferences over (global) possible outcomes of logical dialogues. There remains to identify underlying preferences and strategic inferences that eliminate strategies violating (SR-L) and (SR-R).

The utility function proposed by Hintikka and Hintikka (1983) identifies preferences for a single agent building a systematic proof tree that are grounded in preferences for lower cognitive costs (see Genot 2017). However, reasoning from these preferences cannot be modeled by the standard for strategic inference in extensive games. In a nutshell, a player in an  $n$ -player extensive game is assumed to consider, at any given position where she has to choose an action, all the possible end-states that could result from her next decision, given what she currently knows about the state of Nature and the other players' strategies. Then, she chooses her next move by eliminating moves that belong to strategies yielding lower payoffs, a process called *elimination of dominated strategies*. We have alluded to where the difficulty lies in the single agent case, namely for eliminating dominated proof strategies in the construction of a signed semantic tree (n. 2, p. 6). The problem is essentially the same for the selection of a strategy in a logical dialogue, since both players would have to consider the outcome of infinite strategies in order to eliminate them, which amounts to solving the Halting Problem. Fortunately, there are some other means of strategic inference that are available to players with bounded computational resources, to which we will now turn.

## 4 Preferences and Inferences in Logical Dialogues

### 4.1 Assumptions

Following Barth and Krabbe (1982), we construe logical dialogues as a type of context of rational argumentation, where **P** and **O** have a common goal, namely settle the matter as to whether  $\Gamma$  entails  $\phi$ . They have different stakes, namely **P** is committed to the position that  $\Gamma$  entails  $\phi$ , and **O** is committed the position that it does not. The common goal takes precedence, that is, both prefer to establish something rather than nothing, even at the cost of a revision of their initial position *vis-à-vis*  $\Gamma$  and  $\phi$ . Alternatively, the goal of **P** and **O** can be characterized as *winning* interpreted as *establishing that their initial position vis-à-vis  $\Gamma$  and  $\phi$  is correct*, but with the commitment to acknowledging a loss if the initial position is not correct. There are some difficulties with specifying their goal as winning *simpliciter*, to which we will return in conclusion. Notice that if the matter were whether  $\phi$  is *true*, the argumentation would be best modeled by game-theoretic semantics games, that we discuss in Sect. 6. Below is a set of assumptions that capture logical dialogues so construed.

**Assumption 1** (Common Ground) (1) **P** and **O** know the syntax of  $\mathcal{L}$ , and their action set, that is, their options for attacking and defending sentences of a given syntactic form. (2) **P** and **O** agree that: (a) any sentence  $\gamma \in \Gamma$  is in the common ground; and (b) if  $\beta_1, \dots, \beta_n$ , are sentences in  $\mathcal{L}$ , and are in the common ground, and if  $\beta_m$  is also a sentence in  $\mathcal{L}$ , and must be in the common ground if  $\beta_1, \dots, \beta_n$  are, then  $\beta_m$  is in the common ground; (c) for any  $\beta \in \mathcal{L}$  either  $\beta$  or  $\neg\beta$  must not be in the common ground.

**Assumption 2** (Disagreement) **P** and **O** disagree about  $\phi$ : (1) **P** contends that  $\phi$  is in the common ground as soon as  $\Gamma$  is; but: (2) **O** contends that  $\phi$  need not be in the common ground when  $\Gamma$  is.

**Assumption 3** (Settlement) **P** and **O** both want to know whether  $\phi$  is in the common ground as soon as  $\Gamma$  is, and in particular: (1) prefer to know it as soon as possible; (2) prefer to know rather than not to know, even at the cost of changing their mind; but: (3) otherwise, prefer to hold to their initial opinion as long as they can.

Assumption 1 translates the agreement on  $\mathcal{L}$  and  $\Gamma$  as a precondition for logical dialogue: **1.1** requires agreement on the *particle rules*, irrespective of the grounds of this agreement<sup>8</sup>; **1.2** sets weak constraints on the content of the common ground, in particular that **O**'s defenses against **P**'s attacks are in the common ground. **1.2** expresses the principle of non-contradiction (without commitment to a semantic interpretation). It does not entail that  $\Gamma$  is always consistent, only that  $\Gamma$  *must not* be inconsistent. If it turns out that  $\Gamma$  is inconsistent, then the common ground for the dialogue must be revised. How this revision should be handled is not part of the game.<sup>9</sup>

Assumption 2 characterizes **P**'s and **O**'s epistemic position relative to  $\phi$ , in terms of their initial opinion (guess, conjecture, or whatnot). Together with Assumption 1, 2 entails an asymmetry between **O** and **P**: both **P** and **O** know that **O**'s defenses against **P**'s attacks can induce different sequences of **O**-labeled sentences, and **P** has to show that  $\phi$  is in the common ground relative to *all* these alternative sequences, whereas **O** only has to exhibit one sequence that does not support  $\phi$ . Assumption 3 completes the characterization of **P**'s and **O**'s epistemic position relative to  $\phi$ , and commits them to examine the issue as efficiently as possible (**3.1**), settle it if possible (**3.2**), and not give up without definitive reasons (**3.3**). Assumptions 1–3 do not suffice for either **O** or **P** to form anticipations about each others' moves and their best response to those moves. The following assumption does the job:

**Assumption 4** (Common Knowledge) The content of Assumptions 1–3 is common knowledge between **O** and **P**.

Assumption 4 does not entail that the structure of the game (the complete game tree) is common knowledge unless **O** and **P** have unbounded computational resources. Therefore, even if common knowledge is a strong idealization, Assumption 4 does not sneak in elimination of dominated strategies.

<sup>8</sup> Arguers need to share  $\mathcal{L}$  as a common language for the game to be possible, but need not know more than rules for attacks and defenses, because they suffice to characterize their action set. Alternatively, these rules can be viewed as means to *teach*  $\mathcal{L}$  and its semantics [on this Wittgensteinian interpretation of the rules, see Jacot et al. (2016)]. However, the cardinality of the domain of discourse must somehow be included in the common ground, because it cannot be expressed in  $\mathcal{L}$  (cf. n. 4, p. 7).

<sup>9</sup> A particularly insightful reader could object that ruling out contradictions re-introduces normative considerations, and thus *rules*. To that reader, we respond that we are merely assuming that **O** and **P** agree not to be dialetheists, and thus that we are merely introducing a convention, and modeling players that conform to it.

## 4.2 Strategic Reasoning (I): The Elimination of SR-L

The thesis  $\phi$  is a single finite sentence that **O** can only challenge finitely many times before he is forced to re-iterate some attacks he has already used (possibly with another optional argument). By Assumption 1.1, **P** can expect that if during a given play: (1) **O** targets  $\phi$  with a sequence of attacks and does not repeat any of them; and (2) she defends herself against all these attacks, then there will be a position where her defense must be a literal. By Assumptions 4 and 3.1, **P** expects that **O** will not repeat any attack unless necessary. By Assumption 1 she knows that in order to support her claim that  $\phi$  follows from  $\Gamma$  in that play, it suffices that the literal she has to defend herself with is in the common ground for that play. If furthermore **P** manages to do so for any literal that **O** could ask her in that play, then **O** would have good reasons to also accept  $\phi$ , and by Assumption 3, to change his mind and concede the play. Hence, **P**'s best strategy is to let **O** ask her for literals, and then obtain them from  $\Gamma$ .

By Assumption 4, **O** knows **P**'s preferences and therefore can put himself in her shoes and simulate her strategic reasoning as characterized above. Thus, **O** knows that his best response to **P**'s strategy is to make it as difficult for her as possible to succeed, without preventing the settlement of the issue. In particular, **O**'s best response to **P**, when **P** defends with a literal, is to never state that literal if he has the option not to. By Assumption 4, **P** can also put herself in **O**'s shoes and anticipate **O**'s strategy. Therefore, she can infer that her best strategy is to always *delay* a literal defense until after she has obtained the literal from **O**, unless she can force him to state the literal with an attack for which she can constrain the defense. And again, by Assumption 4, **O** can infer that **P** can infer this much, etc. As a consequence, we have the following:

**Observation 1** (*Strategic Reasoning for Literals*) *The following is common knowledge between **O** and **P**:*

1. ***P**'s best strategy recommends to delay the statement of a literal  $\psi$  in a given play until **O** has stated  $\psi$  in the same play, unless **P** knows that she can force **O** to defend with  $\psi$  at a later stage.*
2. ***O**'s best strategy recommends not to state a literal  $\psi$  if **P** has already stated  $\psi$  in the same play, unless abstaining to do so would delay the settlement of the issue.*

Assuming that **P** always tries to play her best strategy, Observation 1.1 entails that she should in general avoid to state literals unless **O** has already stated them. Therefore, she should behave 'as if' she were complying with (SR-L), in particular when her anticipations of **O**'s moves are limited. Observation 1.2 constrains **O**'s strategies to a lesser extent, essentially due to Assumptions 3.1 and 3.3. Notice that the contribution of Assumptions 3.1 and 3.3. to Observation 1.1 is marginal. If **P** states a literal but fails to obtain it, and then obtains another that she can substitute to the first, she has delayed the settlement of the play by one move. By contrast, **O** has the ability to indefinitely prevent the settlement of the issue by repeating attacks, instead of answering with literals.



### 4.3 Strategic Reasoning (II): Harsanyi Maximin Principle

Observation 1.1 has a straightforward corollary: in the absence of anticipation about **O**'s strategy in a given play, **P**'s best strategy complies with (SR-L) because it is **P**'s *best response to what would be the worst case for her*: a play where **O** could avoid indefinitely stating literal(s) **P** needs. Furthermore, as mentioned above, complying with (SR-L) requires one less move than stating a literal first and substituting it for another, even if **P** can anticipate obtaining the other. Hence, compliance with (SR-L) is a valuable principle for strategy selection, in particular if **P**'s ability to anticipate **O**'s moves is limited. This corollary can be reformulated with a modicum of game-theoretic jargon: in any 2-player non-cooperative game, a strategy for player **X** that is **X**'s best response to **Y**'s strategy most detrimental to **X**, is called **X**'s *maximin strategy*, which yields:

*Remark 2* (Corollary of Observation 1) **P**'s *maximin strategy* never recommends to state a literal  $\psi$  in a given play, unless **O** has already stated  $\psi$  in the same play.

Maximin strategies are central to the theory of 2-player zero-sum games of which logical dialogues are a special case. In those games, the *maximin solution*, which obtains when both players play their maximin strategies, always coincides with a *Nash Equilibrium* solution. Thus, assuming that *both* **P** and **O** should play their maximin strategy in a logical dialogue, we can conclude that a logical dialogue will reach a Nash Equilibrium. We have just seen that the assumption is justified for **P** when **P**'s anticipations about **O**'s moves are limited. In fact, both **P**'s and **O**'s anticipations are limited when we assume that they have bounded cognitive resources. And we have argued following [Hintikka and Hintikka \(1983\)](#) that **O** and **P** should be conceived of as rationally bounded (cf. Sect. 3.2). In that case, choosing among strategic options becomes a problem of strategic decision under uncertainty about the future states of the game. An early result of game theory [proved by [Neumann and Morgenstern \(1944\)](#)] is that every non-cooperative game has a maximin solution. Based on this result, [Harsanyi \(1977\)](#) has proposed that players who reason under uncertainty about each other's strategy in competitive games should comply with the following:

**Definition 3** [Harsanyi Maximin Principle] In a 2-player zero-sum game, if **X** cannot form rational expectations about **Y**'s strategies, then **X** should play her maximin strategy.

In the classical treatment of extensive games, **X**'s rational expectations about **Y**'s strategies are modeled by a (subjective) probability distribution for **X** over **Y**'s strategies. If **X**'s anticipations about **Y**'s strategies are limited, then **X** cannot have a complete representation of **Y**'s strategy space. Therefore, **X** cannot form a representation of the possible histories of the game. Also, **X** cannot distribute meaningful probabilities over a partition of **Y**'s strategies, and update her subjective probability that **Y** is playing this or that strategy, based on the moves she witnesses. Consequently, in a game in extensive form were **X** cannot anticipate all of **Y**'s moves, **X** cannot either ascribe meaningful probabilities to **Y**'s next move. The rationale for the Harsanyi Maximin Principle (HMP) is therefore that, if **X** cannot form rational expectations about **Y**, then

she cannot rationally expect **Y** to play any strategy other than the most harmful for **X**. If we assume that both **O**'s and **P**'s representations of the structure of the game are partial, HMP becomes a natural candidate for guiding their strategic inferences. The other option would be the elimination of dominated strategies, which we have dismissed in Sect. 3.2. Therefore, we have justified the following assumption:

**Assumption 5** (Strategic Anticipations) (1) In the context of a logical dialogue, **P**'s and **O**'s ability to form rational anticipations about each other's strategy is limited. (2) Subsequently, both **P** and **O** select their strategy according to the Harsanyi Maximin Principle.

## 5 Solving Logical Dialogues

### 5.1 Knowing What to Choose

Assumption 1–5 warrant a reformulation of the argument that underlies the first step of the completeness proof for logical dialogues (discussed in Sect. 3.1) from explicit preferences:

**Observation 4** *Under Assumption 1–5, tree-building rules for signed semantic trees express **P**'s and **O**'s preferences for attacks and defenses among options in their action sets. Equivalently, and omitting attacks, **O**'s best moves map to **T**-cases, and **P**'s, to **F**-cases.*

The proof of Observation 4 (given in “Appendix: Proof” section) establishes that maximin strategies under uncertainty for **P** and **O** always recommend options that map on tree-building rules for **F** and **T** cases, respectively. Assuming Assumption 5 guarantees that **P** and **O** play their maximin strategy, and therefore that Observation 4 holds. Although the rationale for **P**'s and **O**'s best options is the same as in the informal arguments, the proof of Observation 4 proceeds from explicit preferences, and is thus the first part of a solution to the BKHH problem.

Observation 4 does not tell the whole story about how **O** and **P** choose their strategies. First, in cases where there is more than one option for attack or defense, Observation 4 makes no recommendation, which amounts to recommending equivocation (lottery with equal weights), although memory of the current play may warrant further recommendations. Assumption 3 entails that **O** and **P** have to engage in more than one play (cf. Sects. 3.1, 3.2) thereby realizing the “strategy games” of Rahman and Keiff (2005). Hence, memory of earlier plays may also warrant additional recommendations. Finally, anticipations about the future of the game, in particular given the preference that both **O** and **P** have for not delaying settlement of the issue at stake (by Assumption 3.1), may also warrant additional recommendations. Since Assumption 3.1 takes care of (SR-R), the only issue remaining for solving the BKHH problem is to reconstruct the reasoning of **O** and **P** about whether or not they should stop a play (and possibly the whole game).

## 5.2 Knowing When to Stop

In a logical dialogue, **O** plays first, and he can either attack  $\phi$  or immediately quit playing, thereby conceding a defeat in the game. Notice that concession of defeat in a play is an action available to players, otherwise termination of a play must be decided by a rule, and the solution to the BKHH problem would be incomplete. By Assumption 2, conceding from the beginning is tantamount to concede that  $\phi$  is in the common ground. But **O** clearly has insufficient grounds to do so in agreement with Assumption 3.3. Therefore, **O**'s best option is to attack. **P**'s options depend on  $\Gamma$ , her anticipations about **O**'s strategy, and how much she can anticipate of her own strategy. If those anticipations are limited, the argument we gave for Observation 1 illustrates which *type* of strategies **P** should implement from that point on. In fact, the less insights she has into **O**'s strategy, the better off she will be letting **O** attack, until a position is reached where she has to defend herself with a literal statement.

After a position is reached where she has to defend with a literal, **P** can in principle implement a strategy whose outcome will be similar to the output of a depth-first systematic method for building tableaux proofs (cf. Sect. 3.2). This follows in particular from Observation 4, which guarantees that the local topology of a tree will always coincide with that of a signed semantic tree. Hence, the only choices that matter are: (1) the order in which the premises in  $\Gamma$  are attacked; and (2) the order of attacks over each premise. Systematic methods for tableaux construction, be they breadth-first as those of Smullyan (1968), or depth-first as those of Rahman and Keiff (2005), prescribe both (1) and (2), and thus are blueprints for systematic strategies for **P**, past the point where she has to defend with a literal.

Some time after **P** has begun to implement a strategy for obtaining literal statements from **O**, a position may be reached where **P** manages to counter **O** using only **O**'s previous statements. Positions of this type are reminiscent of Socrates' typical argumentation strategy, which uses only concessions made by his opponents to support his arguments. Hence, we propose to call them *Socratic Positions*:

**Definition 5** (Socratic Positions) A *Socratic Position* is a position  $m$  in some play of a logical dialogue such that: (a) at some position  $m'$  occurring earlier than  $m$  in the same play, **O** has already attacked every **P**-labeled statement at least once; and (b) at position  $m$ , **P** has defended against all of **O**'s attacks as recommended by her *maximin* strategy.

Socratic positions (hereafter SP) have an important property that makes them of special interest: once a SP is reached in a play, and whatever **O**'s strategy is after the SP is reached, **P** can always force the play to reach another SP (for a proof, see Lemma 10 in "Appendix: Observations 6 and 7" section). Given Definition 3, we have the following:

**Observation 6** *If a Socratic Position is reached in a given play of a logical dialogue, **O**'s maximin strategy recommends to concede the play to **P** and start a new play if possible.*

The proof of Observation 6 (also in "Appendix: Observations 6 and 7" section) is straightforward from the repeatability of SPs. In a nutshell, as soon as a SP is reached,

by Assumption 5 and Definition 3, **O** should behave as if **P** will force another SP. By Assumption 3.1, **O** then prefers to concede the play to **P**, rather than keep playing. By Assumption 3.2, **O** also prefers to concede the game to **P** if he cannot start a new play.

By Assumption 5 and Observation 6, the following is immediate:

**Observation 7** *If a Socratic Position is reached in a given play of a logical dialogue, **O** should concede defeat to **P** in that play.*

By Observation 7, and assuming instrumental rationality, **O** will concede defeat in any play that can be mapped to an atomically closed branch of a signed semantic tree. This takes care of the end-of-play rule for logical dialogues, corresponding to the closure rule for signed semantic trees that closes a branch where  $\mathbf{T}\psi$  and  $\mathbf{F}\psi$  appear together. Thus, there only remains two closure rules, namely those that close a branch in case a contradictory assignment occurs in that branch. By Observation 1.1 and Assumption 5, **P** will never state contradictory literals, unless **O** has. Furthermore, **P** cannot be compelled to state contradictory literals in every play of a game unless  $\phi$  is a contradiction. Hence, if **P** states inconsistent literals, and if  $\phi$  is not a contradiction, there must be some earlier defenses in that play that **P** can revise in order not to be committed to state a contradiction. By Assumption 1.2, contradictions are excluded from the common ground. Therefore, **P** should concede defeat in the play (and indeed the game) as soon as she is compelled to state a contradiction that she cannot avoid. Assumption 3 guarantees that she prefers to know earlier than later whether  $\phi$  is in the common ground, even at the cost of changing her mind, rather than playing indefinitely. The same holds *mutatis mutandis* for **O**. This suffices to establish the following:

**Observation 8** *If a position is reached in a play in which player **X** can be compelled to state a literal that contradicts another literal that **X** has stated previously in the play, then **X** should stop playing and concede defeat to **Y** in that play.*

Observation 8 takes care of the end-of-play rules for classical logical dialogues that force players to stop when a contradiction is reached, and completes the replacement of closure rules by decisions based on preferences.

### 5.3 The Completeness of Logical Dialogues Revisited

The results of the preceding sections establish that players whose preferences and strategic inference processes are characterized by Assumption 1–5 behave ‘as if’ bound by (SR-L), (SR-R), and end-of-play rules that mimic tableaux closure rules. As a consequence, we get:

**Theorem 9** *For a first-order language  $\mathcal{L}$  of arbitrary signature, and any  $\Gamma \subseteq \mathcal{L}$  and  $\phi \in \mathcal{L}$ , the following statements are equivalent:*

1. *There is a closed signed semantic tree with premises  $\Gamma$  and conclusion  $\phi$ .*
2.  *$\Gamma$  entails  $\phi$ —i.e. all the models of  $\Gamma$  are models of  $\phi$ .*
3. *Under Assumptions 1–5, **P** has a winning strategy in a logical dialogue with common ground  $\Gamma$  and thesis  $\phi$ .*

Besides the proviso “Under Assumptions 1–5”, the only difference between Theorem 9 and the completeness theorem for classical logical dialogues given by Rahman and Keiff (2005), is the mention of the common ground  $\Gamma$ . However, Keiff (2009) generalizes the notion of ‘dialogical validity’ to ‘dialogical consequence’, which is the notion we have taken as primitive. Since the BKH problem amounts to finding an alternative account of logical dialogues based on preferences and strategic inferences that would be *equivalent* to the rule-based account, near identity is a feature, not a bug.

## 6 Extending the Solution

### 6.1 Game-Theoretic Semantics

The logical dialogues that we have characterized can be viewed as the resource-bounded realizations of games that idealized players of classical game theory could play with unbounded computational resources. Let us refer to the idealized versions of  $\mathbf{O}$  and  $\mathbf{P}$  as  $\mathbf{O}^*$  and  $\mathbf{P}^*$ . Obviously, Assumption 5 does not hold for  $\mathbf{O}^*$  and  $\mathbf{P}^*$ : they can form anticipations of any depth, and they can reason from a complete representation of the game. For simplicity, we also assume that Assumption 3.1 does not hold for them, so that they are indifferent to the length of a play. Under this last simplification, the strategic form of a logical dialogue with premise  $\Gamma$  and conclusion  $\phi$  (in some language  $\mathcal{L}$ ) is a matrix pairing each of the possible strategies for  $\mathbf{O}^*$  with each of the possible strategies for  $\mathbf{P}^*$ , and where these strategies are characterized as the choice of a *model* for either  $\Gamma$  or  $\phi$ . Payoffs can be set so that  $\mathbf{O}^*$  gets 1 and  $\mathbf{P}^*$   $-1$  when the models are not the same, and conversely  $\mathbf{O}^*$  receives  $-1$  and  $\mathbf{P}^*$  receives 1 when they are the same, where “sameness” is defined as *verifying exactly the same sentences of  $\mathcal{L}$* . Thus,  $\mathbf{P}^*$  has a winning strategy in the strategic game iff, whenever  $\mathbf{O}^*$  chooses a model for  $\Gamma$ ,  $\mathbf{P}^*$  can choose the same model for  $\phi$ .

One can obtain a strategy in the extensive form of the game from a strategy in the strategic form *via* the model-theoretic notion of *complete diagram* of a model. Given a language  $\mathcal{L}$ , the diagram of a model  $\mathfrak{M}$  in which  $\mathcal{L}$  is interpreted, is the (possibly infinite) set of literals formed with all the predicates in  $\mathcal{L}$  (once enough constants have been added to the signature of  $\mathcal{L}$  to name all the elements of the domain of the model) that are *true* in  $\mathfrak{M}$ .<sup>10</sup> A fundamental result of model theory is that two models with the same complete diagram satisfy the same sentences. If  $\mathbf{P}^*$  has a winning strategy in the strategic game, then her winning strategy in the extensive game will always satisfy the constraint set by (SR-L). Indeed, the model that  $\mathbf{O}^*$  and  $\mathbf{P}^*$  choose as a strategy for the game in the strategic form determines the literals  $\mathbf{O}^*$  and  $\mathbf{P}^*$  state, as part of their strategy in the extensive form of the game. But if  $\mathbf{P}^*$  chooses the same model as  $\mathbf{O}^*$ , then she must choose the same complete diagram. Thus  $\mathbf{P}^*$  will always be able to state the same literals as  $\mathbf{O}^*$ .

<sup>10</sup> Note that there may be no enumeration of these constants, if the domain is uncountable. Abstract model theory, where the use of complete diagrams originates, routinely deals with arbitrary signatures, including uncountable sets of constants.

$\mathbf{O}^*$  and  $\mathbf{P}^*$  could always extend a play to the point where both are committed to a fully specified complete diagram for the model they have chosen. For this, players must be allowed to ask binary questions about sentences in  $\mathcal{L}$  (as per the rule of Fig. 5, p. 10). In principle, these questions could be about sentences of arbitrary complexity, but letting them be about atoms of  $\mathcal{L}$  and their negation is sufficient. In order to prove that she has chosen the same model as  $\mathbf{O}^*$ ,  $\mathbf{P}^*$  must then be able to answer all atomic questions in the same way as  $\mathbf{O}^*$  has, that is, with the same literal answer as  $\mathbf{O}^*$ . We can now justify our ‘hybrid’ rules, and why they are necessary to deal with cases where  $\Gamma = \emptyset$  and  $\phi$  is a tautology for  $\mathbf{O}$  and  $\mathbf{P}$ : the strategies of  $\mathbf{O}$  and  $\mathbf{P}$  are not equivalent to *choosing* complete models, but to *building* partial models. When  $\Gamma = \emptyset$ ,  $\mathbf{P}$  has no information about any partial model, and therefore must recover enough information to win a logical dialogue. Furthermore, logical dialogues between  $\mathbf{O}$  and  $\mathbf{P}$ , as we have characterized them, can *in principle* be extended so as to correspond, in the limit of an infinite process, to a dialogue between  $\mathbf{O}^*$  and  $\mathbf{P}^*$ :  $\mathbf{O}$  and  $\mathbf{P}$  can in principle ask each other binary questions about all the atoms of  $\mathcal{L}$ , although they will typically not do so, essentially because of Assumption 3.<sup>11</sup>

‘Hybrid’ logical dialogues are thus ‘model-building’ dialogues, where the partial diagrams are proxies for incompletely specified models. From these games, the model-checking procedure of game-theoretic semantics games is easily obtained. First, we add to the game a third player, *Nature*, whose strategy is the selection of a model  $\mathfrak{M}$ . Second, for a game about  $\phi$ , we set  $\Gamma = \emptyset$  initially, but every time  $\mathbf{P}$  states some  $\neg\psi_i$  as the result of an attack from  $\mathbf{O}$ ,  $\psi_i$  is added to  $\Gamma$ , and is now open to  $\mathbf{P}$ ’s attacks. Similarly, every time  $\mathbf{O}$  states some  $\neg\psi_j$ ,  $\psi_j$  is added to  $\mathbf{P}$ ’s statements, and is open to  $\mathbf{O}$ ’s attacks. Hence, there is, strictly speaking, no common ground, save for the interpretation of  $\mathcal{L}$  in  $\mathfrak{M}$ . Since  $\phi$  is a finite sentence, the game is bound to reach a position where one of  $\mathbf{P}$  or  $\mathbf{O}$  is committed to some atom  $p$  and the other to its negation  $\neg p$ . Then, the player whose turn it is to play can ask a binary question to Nature. If Nature’s answer is the literal that the player who asked the question is committed to, that player wins and the other player loses. Symmetrically, if Nature’s answer is not the literal that the player who asked the question is committed to, that player loses and the other player wins.

Assumption 1–5 must be somewhat reformulated for the above game, but the reformulation amounts essentially to a simplification. Assumption 1.1 remains the same, with the understanding that the players’ action set includes new (mandatory) actions for  $\mathbf{X}$  ( $\mathbf{X} \in \{\mathbf{P}, \mathbf{O}\}$ ) when  $\mathbf{Y}\neg\psi$  occurs at a node. Assumption 1.2 becomes obsolete, as the only common ground is the interpretation of  $\mathcal{L}$  in  $\mathfrak{M}$ . In Assumptions 2 and 3, “is in the common ground” and “is not in the common ground” are sub-

<sup>11</sup>  $\mathbf{O}$  would not use binary questions, because  $\mathbf{P}$  can always adopt a ‘copy-cat’ strategy, and counterattack asking the same question as  $\mathbf{O}$ , and copying his answer. In particular,  $\mathbf{O}$  would *never* extend a play that has reached a SP by asking questions repeatedly, because he would merely delay the settlement of the issue. Consequently, binary questions do not affect Observations 6 and 7. As for  $\mathbf{P}$ , she would not use binary questions as a primary means for obtaining a literal from  $\mathbf{O}$ . Indeed, the worst case for  $\mathbf{P}$  is when  $\mathbf{O}$  states the negation of the literal she needs, and where she is then forced to obtain the literal by attacking earlier statements of  $\mathbf{O}$ , in order to force him to contradict himself. But this strategy alone would give her the literal she needs without the delay of the initial interrogative move. Thus,  $\mathbf{P}$  would prefer not to use interrogative moves in the first place, when  $\Gamma$  is not empty (by Assumption 3.1).

stituted with “is true” and “is false”, respectively. Assumption 2 can be rephrased to refer to the mandatory actions in response to negated statements, characterizing them as an agreement to disagree. Finally, Assumption 4 remains as it is, and so does Assumption 5, although the role of the latter is much less critical. Indeed, the game is simpler, because it involves only attacks on  $\phi$ , its subformulas, and the negation of some of its subformulas. It is easily seen that **P**’s best strategy is to let **O** attack  $\phi$  until a literal is reached and then ask Nature about that literal, and that **P** has a winning strategy in the game iff  $\phi$  is true in  $\mathfrak{M}$  according to the interpretation of  $\mathcal{L}$  in  $\mathfrak{M}$ .

### 6.2 Back into the Fold: Dialogical Orthodoxy

Dialogical logic was initiated in the 1950s by Lorenzen (1958) as a foundation for constructive mathematics and intuitionistic logic, a project that was completed two decades later with Lorenzen and Lorenz (1978). This ‘old school’ of dialogical logic was supplanted by a ‘new wave’ in the 1990s, where dialogical logic became a pluralistic framework for capturing and combining a wide variety of logics. The reader will find a précis of this curious historical turn in Rahman and Keiff (2005), as well as rule sets for the following logics: classical, intuitionistic, free, paraconsistent, connexive, modal (normal and non-normal), hybrid (modal with nominals), and independence-friendly. The new wave combines an ‘everything but the black board’ approach to logical pluralism with a knack for colorful names, illustrated for instance by Frege’s *Nightmare* [Rahman (2001): intuitionistic, paraconsistent and free] or *Dialogic for a Wonderful World* [Rahman (2006): non-normal modal logics with nominals]. There is in principle an instantiation of the BKHH problem for all of the above, and we hope that we will not disappoint our most demanding readers, if we admit that we cannot solve them all. However, we will do our best to solve two special cases, namely vanilla classical dialogues and vanilla intuitionistic dialogues, which differ only by a few structural rules.

The structural rule for repetition has classical and intuitionistic variants, allowing for repetitions under conditions, but their formulations are sometimes inconsistent [Rahman and Keiff (2005) permits ‘strict repetitions’ under classical rules, but Keiff (2009) does not]. Another difference is a closure rule for exchanges (‘rounds’) specifying that under classical rules, **P** can change a past defense but cannot do so under intuitionistic rules. Structural rules often require hosts of auxiliary notions, with sometimes Byzantine differences between different accounts. We need not however wrestle with the details, since solving the BKHH problem can be done by looking only at their effects on game trees. Let us begin with common features of vanilla classical and intuitionistic dialogues. The main difference with the particle rules we have used is that player **X** may incur commitments to formula in addition to  $\Gamma$  (if  $\mathbf{X}=\mathbf{O}$ ) or  $\phi$  (if  $\mathbf{X}=\mathbf{P}$ ). Those commitments are made ‘for the sake of the argument’ and limited to the current play and are: (1) to  $\mathbf{X}\psi_1$ , as an attack against  $\mathbf{Y}\psi_1 \rightarrow \psi_2$ ; and (2) to  $\mathbf{X}\psi$  as an attack against  $\mathbf{Y}\neg\psi$ . As a consequence of (2), (SR-L) is substituted with a rule relative to *atoms*, because the statement of a negative literal  $\mathbf{X}\neg p$  can be attacked with  $\mathbf{Y}p$ . Now, for the differences: with classical rules, a round is open as long as **P** can substitute a **P**-labeled defense to another at any node, provided that she is not

merely delaying the completion of the play. With intuitionistic rules, as soon as **P** has defended herself against some attack, the node that was attacked becomes unavailable for application of *any* rules, making her defense final.

In spite of those differences, our solution to the BKHH problem transfers almost immediately to vanilla classical dialogues, pending minor adjustments. Assumption 1 needs to make explicit that **O** can make temporary additions to the common ground, and Assumption 2 can reflect additions to the action set (as with game-theoretic semantics) as a further elucidation of the disagreement between **O** and **P**; and in Observation 1, Remark 2, and Observation 6 (and the proof of Lemma 10), ‘literal’ must be substituted with ‘atom’. Preferences over options expressed by Observation 4 collapse player-independent options on the equivalent (*modulo* translation) of vanilla signed semantic trees because elementary information is given by atoms. Socratic Positions are obviously stable relative to atoms (positive literals) as a special case of stability relative to literals, which entails Observation 6, completing the correspondence between the branches of game trees for vanilla classical dialogues game trees, and branches of vanilla signed semantic trees. Finally, the strategy profile that solves a given dialogue and realizes **P**’s winning strategy (when there is one) remains the same: it comprises, for **O**, the strategy that recommends attacking **P** $\phi$  and defending against **P**’s attacks when they come without delay; and for **P**, the strategy that lets **O** attack until **O** asks for an atom  $p$ , and then implements some systematic (recursive) procedure to extract  $p$  from  $\Gamma$ .

The solution of the BKHH problem for the orthodox vanilla intuitionistic dialogical logic is deceptively simple. Neither Rahman and Keiff (2005) nor Keiff (2009) propose an explicit justification for the rule sets of intuitionistic dialogues. However, the rule set enforces on **P** the constraints of the Brouwer–Heyting–Kolmogorov (BHK) interpretation of intuitionistic logic. For some formula  $\psi \in \mathcal{L}$ , the BHK imposes the following constraints:

- (1) if  $\psi = (\psi_1 \wedge \psi_2)$ , a proof of  $\psi$  is a pair  $\langle a, b \rangle$  where  $a$  is a proof of  $\psi_1$  and  $b$  is a proof of  $\psi_2$ ;
- (2) if  $\psi = (\psi_1 \vee \psi_2)$ , a proof of  $\psi$  is a pair  $\langle a, b \rangle$  where  $a = 0$  and  $b$  is a proof of  $\psi_1$ , or  $a = 1$  and  $b$  is a proof of  $\psi_2$ ;
- (3) if  $\psi = (\psi_1 \rightarrow \psi_2)$ , a proof of  $\psi$  is a function  $f$  that converts a proof of  $\psi_1$  into a proof of  $\psi_2$ ;
- (4) if  $\psi = \exists x\theta[x]$  (where  $x$  is free in  $\theta$ ), a proof of  $\psi$  is a pair  $\langle a, b \rangle$  where  $a$  is an element of the domain of discourse assigned to  $x$  by some assignation function, and  $b$  is a proof of  $\theta[a]$ ;
- (5) if  $\psi = \forall x\theta[x]$  (where  $x$  is free in  $\theta$ ), a proof of  $\psi$  is a function  $f$  that converts every possible assignation of denotation  $a$  to  $x$  into a proof of  $\theta[a]$ ;
- (6) if  $\psi = \neg\psi$ , a proof of  $\psi$  is a proof of  $\psi \rightarrow \perp$ , where  $\perp$  denotes an arbitrary contradiction (*absurdum*), that is by (3) a function  $f$  that converts a proof of  $\psi$  into a ‘proof’ of  $\perp$ , that is the statement of some contradiction.

Satisfaction of these constraints follows as a consequence of the closure rule for rounds (left to the reader). Subsequently, **P** wins a play if her strategy in that play satisfies the BHK constraints for every defense of every **P** $\psi$  occurring in that play, and **P** has a winning strategy in an intuitionistic dialogue if her strategy satisfies the BHK



constraints in all plays. Since the BHK constraints can apply as a ‘filter’ over proof strategies, an easy and unsubtle solution to the BKHH problem for vanilla intuitionistic dialogues is to make the additional assumption that **P** prefers strategies that comply with those constraints, as part of Assumption 3, making this preference common knowledge by Assumption 4. A more subtle suggestion is to interpret the closure rule for rounds as some form of ‘amnesia’ from **P**’s part, inducing compliance with the BHK constraints without explicit commitment to them. We must admit that this suggestion does not seem to lend itself to any natural interpretation, but it is intriguing and in the same vein as the philosophical speculations that often accompany dialogics.

## 7 Concluding Remarks

We defined and motivated an open problem: characterizing explicit preference profiles and strategy selection as an alternative to sets of rules, relative to both dialogical logic and game-theoretic semantics. As a tribute to the authors who first came to grip with this problem, we called it the *Barth-Krabbe-Hintikka-Hintikka* problem (BKHH). We turned to dialogical logic for the ease of extracting (implicit) preferences from rule sets, and we formulated particle rules defining explicitly action sets, and implicitly model-building games where players build partial diagrams. A play of those games can be extended in principle to a complete diagram, and in the limit coincides with a model (the game coincides in the limit with a model-matching game played by ideal reasoners). We obtained a solution to the BKHH problem for those games, and converted it to solutions for model-checking games of game-theoretic semantics, and for orthodox logical dialogues of dialogical logic, in both their classical and intuitionistic variants.

We will conclude with two remarks. The first is that our solution can in principle be extended to any variation of logic, as far as it can be expressed within the dialogical framework. In this respect, our approach is faithful to the spirit of the new wave of dialogical logic, which sees the approach as a conceptual framework for logical pluralism. Our contribution to this framework is the specification of genuine players in a game-theoretic sense, that is, player endowed with preferences for reaching the end-state of an argumentation. Insofar as the project of dialogical logic can be interpreted as a precise account of formal argumentation (see [Rahman and Keiff 2005](#); [Keiff 2009](#)), our contribution is thus tantamount to specifying formal argumentation *games*, complying with the ‘industry standard’ of game-theory. The conceptual benefit of doing so it that, contrary to the structural rules in orthodox dialogical logic, our assumptions are not normative per se, and do not intend to a priori prescribe rules of play, but describe the preferences of rational agents with bounded cognitive resources engaged in an argumentation procedure. (We did not include rational dialetheists, however.)

Our second remark concerns the difficulties we alluded to in Sect. 4.1 with the notion that arguers are aiming at winning *simpliciter*. The second-best outcome to a win is usually a draw before a loss, and a strict preference for winning over losing may incur a preference *for* delaying tactics to force a draw in an infinite play. Subsequently, Barth and Krabbe acknowledge the need for a higher purpose and suggest that a player should quit playing when their opponent has won by rational means. As a reconstruction of this higher purpose, we proposed that the goal of the player is to learn whether  $\Gamma$

entails  $\phi$ , or (equivalently) to learn if one of the players has a winning strategy, with the understanding that ‘learning’ can mean ‘learning in the limit’, if  $\mathbf{O}$  has only an infinite winning strategy (but see n. 6, p. 11). Whether the commitment to a higher purpose is a consequence of the players’ rationality or an externality is a murky issue, hotly debated among argumentation theorists. It would be presumptuous from our part to claim that we have contributed to illuminate it in any substantial way.

**Acknowledgements** The authors wish to thank Sara L. Uckelman for her substantial comments on a previous version of this article, and for suggestions that greatly improved the overall quality of the final version. All remaining errors are their own.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix: Proofs

**Note:** In all the proofs of this Appendix, we leave implicit the role of the following consequences of Assumptions 1 and 4: that  $\mathbf{O}$  and  $\mathbf{P}$  know that in order to defend that  $\phi$  is in the common ground, it is sufficient to show that all the literal consequences of  $\phi$  are in the common ground; that a literal  $\psi$  is in the common ground if  $\mathbf{O}$  can be forced by  $\mathbf{P}$  to state  $\psi$  after some sequence of attacks, where the first attack targeted some element  $\gamma \in \Gamma$ .

### Observation 4

We give detailed proofs for the unnegated cases of disjunctive and existentially quantified statements, sketch the proofs for their negated cases, leaving the role of our assumptions implicit whenever possible, and leave conjunctive and universally quantified statements to the reader. We omit attacks on double negations, which have no optional argument. Finally, we leave implicit the role of Assumption 2, as it is already factored in Observation 1. The proof is in two parts: the first part shows that  $\mathbf{P}$ ’s and  $\mathbf{O}$ ’s best response to the worst case are equivalent (modulo the mapping of labels) to tree-building rules for signed semantic trees; and the second is an immediate consequence of the first part, and Assumption 5.

*Proof of Observation 4 (Part I) Disjunctions (Defense).* **O-case:** By Observation 1,  $\mathbf{O}$  should avoid conceding sentences from which  $\mathbf{P}$  may later obtain literals from. His best option is thus not to concede either disjunct, and instead counterattack if possible, unless this causes an unnecessary delay (by Assumption 3.1). In that case, his best option is to concede as few sentences as possible. Hence,  $\mathbf{O}$  should never reply with both disjuncts. In the absence of particular insight about  $\mathbf{P}$ ’s future moves, there is no reason for  $\mathbf{O}$  to favor one disjunct over the other. Hence,  $\mathbf{O}$ ’s best response to the worst

case (no possible counterattack, and no insights into **P**'s strategy) is to randomly pick one of the disjuncts.

**P-case:** **P**'s best option is to pick a disjunct that she can ultimately defend using literals conceded by **O**. If she lacks insights about which options will be easier to defend, her best option is to defend with both disjuncts. Given Assumption 1.1, she can later 'opt out' one of them later, if she fails to obtain the literals necessary to defend it, because the meaning of a disjunction only requires that she shows one of them to hold, given the common ground. Hence, **O**'s best response to the worst case (no insights into which disjunct she will later be able to justify, given **O**'s strategy) is to defend with both disjuncts.

*Negated disjunction (Attack).* **O-case:** **O** prefers to minimize length of a play, and is thus better off asking only one (negated) disjunct in any given play. If **P** manages to defend herself, **O** still has the option to concede victory in the current play, and ask for the other (negated) disjunct in the next play. In the absence of insights into **P**'s strategy, there is no reason to favor one disjunct over the other. Hence, **O**'s best response to the worst case (no insight into which disjunct **P** will later be able to use to justify herself) is to pick one (negated) disjunct at random.

**P-case:** In the absence of insights into **O**'s future moves, **P** is always better off trying to obtain as many concessions as possible, in order to increase her options for obtaining literals later in the play. Hence, **P**'s best response to the worst case (no insight into which disjunct she will later be able to use to justify herself) is to pick both disjuncts.

*Existential quantifier (Defense).* **O-case:** If  $\mathcal{L}$  has infinitely many individual names, **O**'s options for asking **P** for an instantiation of an existential statement are nondenumerable (see Fig. 4). If **O** has already attacked some of **P**'s universally quantified statements, and/or if **P** has already been challenged to state a literal, then **O**'s defense will make **P**'s task of stating only literals that **O** has already stated, if his defense introduces individual names that have never been introduced already (by Observation 1). However, if **O** lacks insights about how **P** could use literal statements that **P** could later obtain, and which would feature the individual names that his defense introduces, his best option is to introduce new names one at a time. Moreover, by Assumption 1.1, **O** is allowed to revise his defense, if necessary, by adding new instances—and he will thus chose the option to do so, unless he foresees that such an action would delays the settlement of the issue (by Assumption 3). Hence, **O**'s best response to the worst case (no insight into what use **P** will have for the literal consequences of his defense) is to defend with a single *new* individual name, retaining the option to change that name later.

**P-case:** By Observation 1 **P** should prefer using individual names previously introduced by **O**, which increase her chances to be able to defend the literal consequences of the resulting statement. Moreover, if later in the game, it appears that she cannot do so Assumption 1.1 guarantees that she will be allowed to rephrase her defense, so as to match **O**'s literal statements. Moreover, the fewer instantiations she conceded, the fewer literal consequences she will have to defend. Hence, **P**'s best response to the worst case (no insight into which literal consequences of her defense she will be able

to defend at a later stage) is to defend with a single *old* individual name, retaining the option to change that name later.

*Negated existential quantifier (Attack).* **O-case:** One new individual may be sufficient, if **P** cannot later defend the literal consequences of her defense, and may thus keep the length of the play as short as possible (which is suitable, by Assumption 3.1); moreover, if necessary, **P** retains the option to rephrase his defense (by Assumption 1.1) and to demand for another instantiation, if **P** manages to defend the literal consequences of the first. Hence, **O**'s best response to the worst case (no insight into which literal consequences of her defense **P** will be able to defend at a later stage) is to attack with a single *new* individual name, retaining the option to re-iterate his attack at a later stage.

**P-case:** By Observation 1, **P** will in general prefer using individual names that **O** has used to attack her, as she can use the literal consequences of **O**'s defenses to defend herself later in the game. By the same consideration, how many instantiations **P**'s attack will demand, depend on how many literals different names she has to use for defending the literal consequences of  $\phi$ . While it is not always the case that she is better off asking for only one instantiation, she can always ask for them one at a time, and re-iterate her attacks if necessary (by Assumption 1.1), provided that this does not delay unnecessarily the settlement of a play. In any case, **P** should not introduce new individual names in her attacks, unless she can foresee how she could later use literal consequences of **O**'s defense. Hence, **P**'s best response to the worst case (no insight into which literal consequences of **O**'s defense she will be able to defend herself at a later stage) is to attack with only *old* individual names (but not necessarily only one).<sup>12</sup>  $\square$

*Proof of Observation 4 (Part II)* By Assumption 5, both **O** and **P** lack insights into each other's strategy, and both **O** and **P** select their strategy according to HMP. Hence, they play their maximin strategies. Hence, by Part I of the proof, their best options for attacks and defenses are equivalent to tree-building rules for signed semantic trees, modulo the mapping of **O**'s moves (attacked statements and their defenses, omitting **P**'s attacks) to **T**-cases, and the mapping of **P**'s moves (similarly omitting **O**'s attacks) to **F**-cases.

## Observations 6 and 7

We first prove Observation 6, and then Observation 7. To prove Observation 6, we first prove the following lemma:

**Lemma 10** (*Stability of SPs*) *If, in a given play, a SP has been reached, then **P** can force any extension of that play to reach a new SP.*

<sup>12</sup> Notice that in the case **P** asks for multiple instantiations, the number of her options is bounded by the number of individual names that **O** has introduced in the past history of the game. Assuming that **O** plays his maximin strategy, the number of options for **O** will be, at any given time, finite (since the power set of a finite set has finitely many elements). Hence, the game tree generated by **P** and **O**, when they play their maximin strategies, is *finitely generated*, even if the game does not have in principle a finite horizon.

The proof of Lemma 10 is by induction on the number of attacks before a SP that **O** could repeat without changing play. For the induction clause, we assume that for the  $n$  last attacks before the SP has been reached, and that **O** could repeat without changing play, **P** can force the extension of the play to reach a new SP; and we show that if **O** repeats the  $(n + 1)$ -th last attack, **P** will be able to reach a new SP. We give a detailed proof for the base step, and sketch the induction step.

*Proof of Lemma 10 (Base Step)* For the base clause, the proof is by cases, for the *last* attack before the SP is reached. For all the cases, we assume that a SP has been reached, and consider what the last attack was. We divide the case in two types: operators, and quantifiers.

**Case Ia: P-labeled conjunction.** If **O** repeats the attack asking for a different conjunct, **O** is *de facto* generating a new play, and there is nothing to prove.<sup>13</sup>

If **O** repeats the attack asking for the same conjunct, **P** can simply repeat her defense. *Ex hypothesis*, this resulted in a SP the first time. Therefore, the play will reach a new SP.

**Case Ib: P-labeled negated disjunction.** Identical to **Case Ia**, substituting “**P**-labeled conjunction” with “**P**-labeled negated disjunction” and “conjunction” with “negated disjunct”.

**Case Ic: P-labeled disjunction.** If **O** repeats the attack, **P** can pick the same disjunct as the first time. *Ex hypothesis*, this resulted in a SP the first time. Therefore, the play will reach a new SP.

**Case Id: P-labeled negated conjunction.** Identical to **Case Ic**, substituting “**P**-labeled disjunction” with “**P**-labeled negated conjunction” and “disjunct” with “negated conjunct”.

**Case IIa: P-labeled universal statement.** If **O** attacks the **P**-labeled existential statement occurring earlier than the SP with the same individual name, **O** can repeat her defense. *Ex hypothesis*, this resulted in a SP the first time. Therefore, the play will reach a new SP. If **O** repeats the attack with a different individual name, then **P** can repeat the sequence of moves she has made, that lead to the SP, substituting the new individual name to the one used by **O** the first time around whenever necessary. *Ex hypothesis*, this sequence of moves resulted in a SP the first time for the literals mentioning the individual name introduced by **O** the first time. Hence, it will also result into a SP the second time, for all the literals mentioning the individual name introduced by **O** the second time.

**Case IIb: P-labeled negated existential statement.** Identical to **Case Ia**, substituting “**P**-labeled existential statement” with “**P**-labeled negated universal statement”.

<sup>13</sup> The (‘non-strict’) repetition of an attacks, on a **P**-labeled conjunction (or a **P**-labeled negated disjunctions) asking for the other conjunct (negated disjunct) is tantamount for **O** to: (a) conceding the current play and closing it; and: (b) starting a new play in which all the moves of **O** and **P** are repeated until the position is reached where **O** specifies the new option for the attack.

**Case IIc: P-labeled existential statement.** If **O** repeats the attack, **P** can pick the same individual name for stating her defense as she did the first time. *Ex hypothesis*, this resulted in a SP the first time. Therefore, the play will reach a new SP.

**Case IIId: P-labeled negated universal statement.** Same argument as **Case IIc**.

Since cases (I–II) cover all the cases where **O** can repeat his last attack in a play before the play reaches a SP, and since in each of those cases, if the play has reached a SP, then **P** can play so as to reach a new SP, then if **O** repeats the *last* attack before a play has reached a SP, then **P** can force any extension of the play to reach a new SP.

*Proof of Lemma 10 (Induction step, sketch)* **Induction hypothesis:** For the  $n$  last attacks made by **O** before the SP has been reached that **O** could repeat, without switching play, **P** can force the extension of the play that obtains when **O** repeats any of those attacks, to reach a new SP.

A complete proof would be by cases, but all cases have the same structure: for each possible attack for the  $(n + 1)$ -th attack, if **O** repeats the attack and does not switch play, **P** can repeat her defense (possibly substituting a new individual name for the one used the first time around). From that point, she can simply repeat the moves she played the first time around (up to a possible substitution of individual name). Since, *ex hypothesis*, none of the changes that **O** could make in his attacks between the  $(n + 1)$ -th attack and the SP would prevent **P** to reach a new SP, if **O** repeats the  $(n + 1)$ -th attack before a play has reached a SP, then **P** can force any extension of the play to reach a new SP.

Since: (1) **O** cannot prevent a play that has reached a SP by repeating the last attack he made before the play reached a SP; (2) if **O** cannot prevent a play that has reached a SP by repeating the  $n$ -th last attack he made before the play reached a SP, then **O** cannot prevent that play to reach a SP by repeating the  $(n + 1)$  last attack he made before the play reached a SP; and: (3) there is a first attack in the play; then **O** cannot prevent a play that has reached a SP to reach a new SP. Conversely, if a play has reached a SP, then **P** can force any extension of that play to reach a new SP.

We now prove Observation 6.

*Proof of Observation 6* By Lemma 10, if **O** attempts to extend a play that has reached a SP, and if **P** plays the most harmful strategy for **O**—namely, forcing the play to reach a new SP—**O** cannot improve his prospect for victory in play by extending it. By Assumption 3, **O** prefers to know earlier than later, including when he is constrained to change his mind. Therefore, upon reaching a SP, **O**'s best response to the worst case (the strategy whereby **P** forces the play to reach a new SP) does not recommend to extend the current play, but instead to concede defeat in that play.

Finally, we prove Observation 7.

*Proof of Observation 7* By Observation 6, **O**'s maximin strategy never recommends to extend a play where a SP has been reached, but to concede defeat in that play. By Assumption 5.2, **O** should play his maximin strategy once a SP is reached. Therefore, once a SP is reached, **O** should concede defeat in the play.

## References

- Barth, E. M., & Krabbe, E. C. (1982). *From axiom to dialogue: A philosophical study of logics and argumentation*. Berlin: Walter de Gruyter.
- Beth, E. (1955). Semantic entailment and formal derivability. *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen*, 18, 309–342.
- Genot, G. E. (2017). Strategies of inquiry. *Synthese*. doi:10.1007/s11229-017-1319-x.
- Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.
- Hintikka, J., & Hintikka, M. B. (1983). The sign of three: Peirce, Dupin, Holmes. In U. Eco & T. A. Sebeok (Eds.), *Sherlock Holmes confronts modern logic: Toward a theory of information-seeking through questioning*. Bloomington: Indiana University Press.
- Hintikka, J., & Sandu, G. (1997). Game-theoretical semantics. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 361–410). Amsterdam: Elsevier.
- Jacot, J., Genot, E., & Zenker, F. (2016). From reasonable preferences, via argumentation, to logic. *Journal of Applied Logic*, 18, 105–128.
- Keiff, L. (2009). Dialogical logic. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. (Summer 2009 edition).
- Kelly, K. T. (2004). Uncomputability: The problem of induction internalized. *Theoretical Computer Science*, 317(1–3), 227–249.
- Lorenzen, P. (1958). Logik und agon. *Atti del XII Congresso Internazionale di Filosofia*, 4, 187–194.
- Lorenzen, P., & Lorenz, K. (1978). *Dialogische Logik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Rahman, S. (2001). On Frege's nightmare: Ways to combine paraconsistent and intuitionistic free logic. In H. Wansing (Ed.), *Essays on non-classical logic, chapter 3* (pp. 61–85). London: World Scientific.
- Rahman, S. (2006). Non-normal dialogic for a wonderful world and more. In J. van Benthem, G. Heinzmann, M. Rebuschi, & H. Visser (Eds.), *The age of alternative logics: Assessing philosophy of logic and mathematics today, chapter 20* (pp. 311–334). Netherlands: Springer.
- Rahman, S., & Keiff, L. (2005). Logic, thought and action, LEUS. In D. Vanderveken (Ed.), *On how to be a dialogician*. Dordrecht: Kluwer.
- Smullyan, R. (1968). *First-order logic*. New York: Springer.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.