Article submitted to Nordic Journal of Philosophical Logic

Neil Tennant

Game Theory and Convention T

Mail address: Department of Philosophy
The Ohio State University
Columbus, Ohio 43212

E-mail: tennant.9@osu.edu
Telephone: +1-(614)292-1591
Telefax: +1-(614)292-7502

Department: Department of Philosophy
Institution: The Ohio State University

Last revision: September 20, 2004

Short title (opt.): Game Theory and Convention T

Keywords: Game theory; semantics; axiom of choice; Convention T; truth; meaning; translation

Printed September 20, 2004

Neil Tennant

Game Theory and Convention T

Abstract

This paper rebuts criticisms by Hintikka of the author's account of game-theoretic semantics for classical logic. At issue are (i) the role of the axiom of choice in proving the equivalence of the game-theoretic account with the standard truth-theoretic account; (ii) the alleged need for quantification over strategies when providing a game-theoretic semantics; and (iii) the role of Tarski's Convention T. As a result of the ideas marshalled in response to Hintikka, the author puts forward a new conjecture concerning the relationship among truth, meaning and translation.

1 Introduction

This paper answers some criticisms raised by Jaakko Hintikka (2000) against my treatment (Tennant, 1978) of game-theoretical semantics for classical first-order languages. In order to answer these criticisms, I re-describe here the kind of game, and the kinds of objects on which it is played, in a semantically non-committal way. This enables one to formulate an interesting philosophical conjecture about the connections among truth, translation, and Tarski's Convention T. To anticipate: suitable syntactic constraints on a mapping τ from object-language to metalanguage ensure that the provability of all instances of the adequacy schema

$$\mathcal{T}(\varphi) \leftrightarrow \tau(\varphi)$$

suffices to determine both that $\mathcal{T}$ is the truth-predicate for the object language and that $\tau(\varphi)$ is the translation of φ into the metalanguage. That is, one does not need to begin with one or other of these concepts (truth or translation) and then rely on Convention T to determine the other one (translation or truth, respectively). Rather, one can attain both concepts 'simultaneously' (and uniquely) as the solutions to the infinite system of logical 'equations' provided by the Tarskian biconditionals.

I am grateful to Hintikka for providing the initial impetus to look at these matters more closely. It has made the task of rebuttal more fruitful, philosophically, than it would have been had it been limited merely to correcting misinterpretations. I turn now to deal with his criticisms, and to work toward the conjecture just described.

2 Background

It was perhaps my review essay (Tennant 1998—henceforth: *Games*) of Hintikka 1996 (henceforth: *PMR*) that prompted what might strike the reader as the rather more personal parts of Hintikka's reply.[1]

The overall assessment in *Games* was that *PMR* contained 'logical errors, oversights, misjudgments, and/or failures to anticipate objections.' This assessment is still warranted. *Games* went to unusual lengths to give a faithful summary, chapter by chapter, of what the book set out to accomplish. It then entered many substantial criticisms, organized under twenty two different topic headings. Hintikka has not yet addressed any of the deficiencies appearing under these various headings.

Hintikka's focus in his reply is on my own treatment of game-theoretical semantics in *Natural Logic* and specifically with what he takes me to have claimed (in *Games*) about the axiom of choice in that connection. Hintikka writes (p. 134):

> In his review article [*Games*] ... [Tennant] claims that in ordinary ("dependence handicapped") first-order languages it is possible to prove the equivalence of game-theoretical and conventional truth definitions with the axiom of choice.

In *Games* I made no such claim. Hintikka provides no page reference, and no quotation, to back this attribution. I rather suspect that the word 'with' is a typo, and should read 'without'. For at p. 113 of *Games* I had quoted the following claim by Hintikka on p. 32 of his book:

> ... as far as first-order logic is concerned, the game-theoretical truth definition ... is equivalent to the usual Tarski-type truth definition, assuming the axiom of choice.

[1] *Loc. cit.*

I then went on to write (pp. 113–4)

> This equivalence does not actually need the axiom of choice for its proof; for the notion 'player T (or player F) possesses a winning strategy in such-and-such state of play' *can be rigorously defined without any explicit quantification over functions.* (My emphasis here.)

It should therefore have been clear from the context that I was considering a game-theoretical definition of truth different from that for which Hintikka now usurps the honorific title 'the one used in GTS'. People other than Hintikka work in what may broadly be called game-theoretical semantics; and their methodologies may interestingly differ. There is no reason (*pace* Hintikka) why a game-theoretical semanticist has to resort to quantifying over strategies while yet giving a perfectly game-theoretical semantics for a first-order language.

Indeed, such quantification over strategies became an *ideé fixe* of Hintikka only some years after his initial investigations of games for quantifiers, in Hintikka 1973. In that book there was no mention of partially ordered quantifier prefixes, and no talk about independence-friendly languages. Instead, Hintikka offered the game-theoretical approach as illuminating, in a Wittgensteinian way, the linguistic devices of ordinary universal and existential quantification. Talk of seeking and finding individuals struck him as particularly apt when dealing with the quantifiers. These 'moves' took place in an imaginary language game representing a dialectical confrontation between Myself and Nature. Hintikka called these *material* games, by way of contrast with the *formal* games of Lorenzen and Lorenz 1978. In the latter games, winning strategies correlated with proofs of logical validity. There was no assumed model in the background, providing an interpretive backdrop to the play. But in the former games (the 'material' ones), the players contested the truth or falsity of a sentence against precisely such a backdrop. They were disputing the truth-value of the sentence in the model in question—not its logical validity.

3 Game-theoretical semantics for classical logic

In *Natural Logic* I re-described Hintikka's material game between two players. I had them occupying roles T and F (suggestive of truth and falsity). They were contesting the claim made by a sentence φ (taken already to be an object admitting of semantical interpretation); and play would take place against the background of a model M (providing the context for such interpretation). In other words, in the exposition itself I was already describing matters in semantically laden terms.

I stated the theorem that player T possesses a winning strategy on φ in M iff φ is true in M. The inductive proof of this result was obvious, because of the thoroughgoing homology that had previously been secured between the inductive definition of Tarskian satisfaction and the inductive definition of the game-theoretical notion 'possessor of a winning strategy at such-and-such state of play' (see below).

Hintikka says that my own game-theoretical definition of truth in *Natural Logic*

> never uses [the game-theoretical idea of defining truth as the existence of winning strategy (*sic*)] or any other genuinely game-theoretical concept. He does not even have symbols for strategies, let alone quantification over them.[2]

But the definition did not have to quantify over strategies! All that a winning strategy does is tell the player what simpler game (or 'state of play') to opt for at the next stage. The treatment in *Natural Logic* was intended precisely to have the *virtue* that no appeal to the axiom of choice would be needed in order to correlate the truth (in an arbitrary model M) of a sentence φ with player T's possession of a winning strategy on φ in M. Hintikka writes (p. 134)

> [Tennant's] truth conditions for quantified sentences are old-fashioned compositional truth-conditions which have nothing to do with game-theoretical semantics except for the fact that Tennant uses the vocabulary of games in formulating them.

Is this a fair assessment? I shall set out below essentially the same ideas again, but this time without any semantical presuppositions as I set up the game and the theory governing it. By thus separating purely syntactic and game-theoretical considerations from semantical ones, I shall show that Hintikka has missed the philosophical subtlety behind this approach to game-theoretical semantics for classical first-order languages. (This is indeed a pity, since it was Hintikka who

[2] *Loc. cit.*, p. 134.

first proposed game-theoretical semantics as an alternative to the usual Tarskian semantics for ordinary first-order languages with the standard quantifiers.)

I shall describe the game in as neutral a way as possible, so as not to suggest that there is anything semantical about it. The game-theory will be fully developed before any question of semantical interpretation arises. In this way, the equivalence between a certain game-theoretic condition on 'play-charts', and the *truth-conditions* of those same objects *once they are thought of as sentences making assertions* will be all the more illuminating. Moreover, we shall see that we *still* do not need to quantify over strategies in order to cast this game-theoretical light on matters of semantic interpretation. That is, the game-theoretical insights are attained at no ontological costs beyond those incurred originally by Tarski in stating his theory of satisfaction and truth for first-order languages.

4 The game, described asemantically

Let there be two players, α and β. There are also two 'dunce hats', labelled, respectively, 1 and 2. At any stage of the game, each player will have exactly one of these hats on her head. Let the 1-1 assignment of hats to players be called R. Thus if $R(1) = \alpha$ (whence $R(2) = \beta$), player α has hat 1 on her head, while player β has hat 2 on hers. If the players were to exchange their hats, the new hat-assignment would be $\overline{R}$. Hence $\overline{\overline{R}} = R$.

The players are going to play the game with respect to 'play charts' and a 'playing-field'. A playing-field is a collection of individuals, each with various 'properties' and perhaps also standing in various 'relations'. These properties and relations will be represented by predicate-letters (or simply, predicates). If one wanted to make a 'playing-field' for the purposes of having friends over to play the game, one might use a collection of solid shapes of different colors and sizes. One-place predicates could then represent shapes and colors, and two-place predicates such relations as 'is bigger than', 'is darker than', etc. If P is an n-place predicate, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are individuals in the playing-field M, then we shall write $\underline{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ when those individuals stand in the relation represented by P in the playing-field M. The statement $\underline{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ belongs to the formalized language in which we are giving the game theory. Note that we do not yet call this formalized language a metalanguage!—for there is, as yet, no 'object language' under consideration.

The play-charts to be used alongside the playing-field are constructed by an inductive process. The basic materials from which one constructs play-charts are tags $x, y, z \ldots$; predicate-letters; and the three numerals 0, 1 and 2. All these primitive constituents (and those to be introduced below) are assumed to be distinct. That is, there is no overlap among any pairs of categories of primitive constituents. This is a standard condition to impose upon inductively defined entities, in order to secure the uniqueness of their method of composition. It is not peculiar to the study of semantically interpretable syntactic complexes.

First there are the *basic* play-charts. These are formed by concatenating one predicate letter and an appropriate number of tags (among which repetitions may be permitted). Thus, with the two-place predicate L, one could form such basic play-charts as Lxy, Lxx, etc. The tags in a basic play-chart are said to be *untied*.

Secondly there are the 'composite' play-charts. If φ and ψ are any two play-charts, the following will also be play-charts:

- 0φ (whose untied occurences of tags will be those of φ);
- $1\varphi\psi$ and $2\varphi\psi$ (whose untied occurrences of tags will be those that are untied in φ and those that are untied in ψ).

Thirdly there are the 'tying' play-charts. Let φ be a play-chart with an untied occurrence of x. Then $1x\varphi$ will be a play-chart, whose untied occurrences of tags will be those of φ, except for the untied occurrences of the tag x in φ, all of which occurrences are now said to be *tied* by the prefix $1x$. Similarly for 2 in place of 1.

A play-chart with no tags untied is called a *starting play-chart.*

A play of the game involves a starting play-chart φ and a playing-field M. The two players will start by tossing a coin. The winner of the toss will choose which hat (1 or 2) to wear at the outset. That choice of course determines the initial hat-assignment R for the play of the game. Play will consist of finitely many successive stages. Each stage will be characterized by (i) the play-chart ψ currently attended to; (ii) the current hat-assignment; and (iii) which individuals in M have thus far been labelled by which tags untied in ψ (see the rules below). Thus a stage (or state of play) against the background of M can be represented as

$$[\psi, R, f],$$

where f is a finite mapping of the untied tags of ψ to individuals in M. The notation $f(x/\mathbf{x})$ will represent the assignment that results from f by extending it or modifying it so that the tag x is assigned the individual $\mathbf{x}$.

With a starting play-chart φ (which has no untied tags) the initial state of play will be of the form

$$[\varphi, R, \emptyset].$$

where $\emptyset$ is the empty set. The rules of the game are as follows:

1. If the current state of play is of the form $[0\psi, R, f]$ then the players must advance to the state of play $[\psi, \overline{R}, f]$. They have no choice in the matter. (The effect of the prefix 0 is thus to make the players *swap hats* as player 1 and player 2.)

2. If the current state of play is of the form $[n\psi\theta, R, f]$ ($n = 1$ or 2), then $R(n)$ chooses which of the following states of play will indeed be the next one: $[\psi, R, f]$; or $[\theta, R, f]$.

3. If the current state of play is of the form $[nx\psi, R, f]$ ($n = 1$ or 2), then $R(n)$ chooses an individual $\mathbf{x}$ from the playing-field M, and the next state of play will be $[\psi, R, f(x/\mathbf{x})]$.

4. If the current state of play is of the form $[P(x_1, \dots\ x_n), R, f]$, then play stops. If $\underline{P}(f(x_1), \dots, f(x_n))$, then $R(1)$ has won; otherwise, $R(2)$ has won.

Given these constitutive rules for this two-person, zero-sum game, it is easy to define the notion

$$\mathcal{P}_M[\psi, R, f]$$

(the player who can win, in state of play $[\psi, R, f]$, against the background of playing-field M). First we consider the two cases of states of play in which the players do not exercise any choices. The degenerate case where play ends (in a state of play involving a basic play-chart) is easy: the player who *can* win is simply the player who *does* win. Thus

$$\mathcal{P}_M[P(x_1, \dots, x_n), R, f] = R(1) \text{ if } \underline{P}(f(x_1), \dots f(x_n));$$

and

$$\mathcal{P}_M[P(x_1, \dots, x_n), R, f] = R(2) \text{ if not-}\underline{P}(f(x_1), \dots f(x_n)).$$

Another easy case is where the state of play is of the form $[0\psi, R, f]$. Since the players advance to the next state of play without exercising any choices, the *person* who can win immediately before the hat-swap is the one who can win immediately after the hat-swap. Thus

$$\mathcal{P}_M[0\psi, R, f] = \mathcal{P}_M[\psi, \overline{R}, f].$$

Or, put another way,

$$\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff } \mathcal{P}_M[\psi, \overline{R}, f] = \overline{R}(2)$$

—whence also

$$\mathcal{P}_M[0\psi, R, f] = R(2) \text{ iff } \mathcal{P}_M[\psi, \overline{R}, f] = \overline{R}(1).$$

Next we consider states of play in which the players do exercise choices. First we look at such states of play involving composite play-charts. What happens in a state of play of the form $[n\psi\theta, R, f]$ ($n = 1$ or 2)? The player $R(n)$ gets to choose which of ψ or θ to have in the next state of play. That is, $R(n)$ gets to determine whether the next state of play is $[\psi, R, f]$ or $[\theta, R, f]$. Now here is a simple conceptual point about being in a winning position: *when it is your turn to make a choice, you are in a winning position before your choice if and only if there is a choice open to you that puts you in a winning position after it.* Moreover, *if (and only if) at any stage all your choices would put your opponent in a winning position, then at that stage your opponent is in a winning position.* Thus

$$\mathcal{P}_M[1\psi\theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1);$$

and

$$\mathcal{P}_M[1\psi\theta, R, f] = R(2) \text{ iff } \mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2).$$

Likewise

$$\mathcal{P}_M[2\psi\theta, R, f] = R(2) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(2) \text{ or } \mathcal{P}_M[\theta, R, f] = R(2);$$

and

$$\mathcal{P}_M[2\psi\theta, R, f] = R(1) \text{ iff } \mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1).$$

Finally, we look at states of play involving tying play-charts. What happens in a state of play of the form $[nx\psi, R, f]$ ($n = 1$ or 2)? The player $R(n)$ gets to choose an individual $\mathbf{x}$ from the playing-field, and to tag it as x. Thus $R(n)$ gets to determine the next state of play, which has to be of the form $[\psi, R, f(x/\mathbf{x})]$. The simple conceptual point reasserts itself: *when it is your turn to make a choice, you are in a winning position before your choice if and only if there is a choice open to you that puts you in a winning position after it.* Moreover, *if (and only if) at any stage all your choices would put your opponent in a winning position, then at that stage your opponent is in a winning position.* Thus

$$\mathcal{P}_M[1x\psi, R, f] = R(1) \text{ iff for some } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(1);$$

and

$$\mathcal{P}_M[1x\psi, R, f] = R(2) \text{ iff for every } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(2).$$

Likewise,

$$\mathcal{P}_M[2x\psi, R, f] = R(2) \text{ iff for some } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(2);$$

and

$$\mathcal{P}_M[2x\psi, R, f] = R(1) \text{ iff for every } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(1).$$

5 At every state of play, exactly one player is in a winning position

Lemma. For all distinct α, β and for all 1-1 functions $R : \{1, 2\} \to \{\alpha, \beta\}$ the inductively defined function $\mathcal{P}$ from states of play to $\{\alpha, \beta\}$ $(= \{R(1), R(2)\})$ is total.

Proof. By induction on the structure of play-charts.

Basis. On states of play involving basic play-charts, it is determinate whether or not the state of affairs $\underline{P}(f(x_1), \ldots f(x_n))$ obtains. So $\mathcal{P}$ is total on these states of play.

Inductive Hypothesis. Assume that $\mathcal{P}$ is total on all states of play involving simpler play-charts than the play-chart φ under consideration. We consider φ by cases, according to whether it is of the form (i) 0ψ, (ii) $1\psi\theta$, (iii) $2\psi\theta$, (iv) $1x\psi$ or (v) $2x\psi$ (where ψ and θ are of course simpler than φ).

Case (i): φ is 0ψ. By definition,

$$\mathcal{P}_M[0\psi, R, f] = \mathcal{P}_M[\psi, \overline{R}, f].$$

By Inductive Hypothesis, the right-hand side exists; whence the left-hand side $\mathcal{P}_M[0\psi, R, f]$ is well-defined.

Case (ii): φ is $1\psi\theta$. The definition of $\mathcal{P}$ specifies that

$$\mathcal{P}_M[1\psi\theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1).$$

By Inductive Hypothesis,

$$\text{either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\psi, R, f] = R(2);$$

and

$$\text{either } \mathcal{P}_M[\theta, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(2).$$

This yields the following four cases to consider:

$$\mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1);$$

$$\mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2);$$

$$\mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1);$$

$$\mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2).$$

In the first three of these cases, it follows by definition that

$$\mathcal{P}_M[1\psi\theta, R, f] = R(1).$$

In the fourth case, it follows by definition that

$$\mathcal{P}_M[1\psi\theta, R, f] = R(2).$$

Hence $\mathcal{P}_M[1\psi\theta, R, f]$ is well-defined.
Case (iii) is similar to case (ii).
Case (iv): φ is $1x\psi$. The definition of $\mathcal{P}$ specifies that

$$\mathcal{P}_M[1x\psi, R, f] = R(1) \text{ iff for some } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(1);$$

and

$$\mathcal{P}_M[1x\psi, R, f] = R(2) \text{ iff for every } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(2).$$

Unlike Case (ii), we cannot give a finite proof by cases here, for there might be infinitely many individuals $\mathbf{x}$. Note, however, that by Inductive Hypothesis we have

$$\text{for every } \mathbf{x} \text{ in } M \text{ either } \mathcal{P}[\psi, R, f(x/\mathbf{x})] = R(1) \text{ or } \mathcal{P}[\psi, R, f(x/\mathbf{x})] = R(2).$$

This is of the form $\forall x(Ax \vee Bx)$. From this it follows by classical logic that $\exists x Ax \vee \forall x Bx$. But the first disjunct is the condition for

$$\mathcal{P}_M[1x\psi, R, f] = R(1)$$

to hold; while the second disjunct is the condition for

$$\mathcal{P}_M[1x\psi, R, f] = R(2)$$

to hold. It follows that $\mathcal{P}_M[1x\psi, R, f]$ is well-defined.
Case (v) is similar to case (iv). *QED*

6 The most important biconditionals; and a homomorphism

Equipped with this Lemma, we can now deduce certain consequences of our definitional clauses above.

Making the classical assumption that it is determinate whether or not any basic fact obtains, we can infer:

$$\boxed{\mathcal{P}_M[P(x_1, \ldots, x_n), R, f] = R(1) \text{ if and only if } \underline{P}(f(x_1), \ldots f(x_n))}$$

Because $\mathcal{P}$ is total, the clause for hat-swaps implies:

$$\boxed{\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff not-}[\mathcal{P}_M[\psi, \overline{R}, f] = \overline{R}(1)]}$$

Recall also that we already have the following:

$$\boxed{\mathcal{P}_M[1\psi\theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1)}$$

$$\boxed{\mathcal{P}_M[2\psi\theta, R, f] = R(1) \text{ iff } \mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1)}$$

$$\boxed{\mathcal{P}_M[1x\psi, R, f] = R(1) \text{ iff for some } \mathbf{x} \text{ in } M \ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(1)}$$

$$\boxed{\mathcal{P}_M[2x\psi, R, f] = R(1) \text{ iff for every } \mathbf{x} \text{ in } M\ \mathcal{P}_M[\psi, R, f(x/\mathbf{x})] = R(1)}$$

Thus far we have not sought to interpret play-charts as making statements about the playing-field. Let φ be a play-chart and let f assign to each tag untied in φ either an individual (from M) or a variable in our formalized language for game theory. No two distinct tags may be assigned the same such variable by f.[3] Suppose f is an assignment dealing with exactly the untied tags in φ, and suppose further than ψ is a subformula of φ. Then f^{ψ} will be the restriction of f to tags that are free in ψ. We now define a homomorphic mapping τ_M^f from play-charts φ, and relative to such assignments f, to sentences of our formalized language of game theory.

$$\tau_M^f(P(x_1, \ldots x_n)) =_{df} \underline{P}(\tau_M^f(x_1), \ldots \tau_M^f(x_n));$$

$$\tau_M^f(0\psi) =_{df} \neg\tau_M^f(\psi);$$

$$\tau_M^f(1\psi\theta) =_{df} (\tau_M^{f^{\psi}}(\psi) \vee \tau_M^{f^{\theta}}(\theta));$$

$$\tau_M^f(2\psi\theta) =_{df} (\tau_M^{f^{\psi}}(\psi) \wedge \tau_M^{f^{\theta}}(\theta));$$

$$\tau_M^f(1x\psi) =_{df} \exists\mathbf{x}\tau_M^{f(x/\mathbf{x})}(\psi);$$

$$\tau_M^f(2x\psi) =_{df} \forall\mathbf{x}\tau_M^{f(x/\mathbf{x})}(\psi).$$

Note that we have defined nothing more than a syntactic mapping, from play-charts to sentences of the formalized language in which we are presenting our game theory.

Here is an example to illustrate the action of τ_M^f as just defined. Consider the play-chart

$$2y10Ayx1xBx.$$

[3] Strictly speaking, since this discussion of f and τ_M^f is taking place a language one level up from the language of our game theory, we should use, as the mapping referred to by this superscript, not f itself, but a mapping f^* closely related to f. Where f assigned to any tag an individual from M, f^* would assign to that tag some name (in the language of our game theory) for the individual. This would ensure that $\tau_M^{f^*}(\varphi)$ would always be a sentence of the language of our game theory. Here, however, we simplify slightly by avoiding such niceties. Alternatively, we could just stipulate that the language of our game theory simply contained all the relevant expressions of its own metalanguage.

Note that the tag x has both tied and untied occurrences therein. The leftmost occurrence of x is untied. No other tag has an untied occurrence. So suppose f maps x to the individual γ in M. Then f is simply $\{< x, \gamma >\}$. We can now calculate $\tau_M^{\{<x,\gamma>\}}(2y10Ayx1xBx)$ as follows:

$$\tau_M^{\{<x,\gamma>\}}(2y10Ayx1xBx)$$

$$= \forall \mathbf{y} \tau_M^{\{<x,\gamma><y,\mathbf{y}>\}}(10Ayx1xBx)$$

$$= \forall \mathbf{y}(\tau_M^{\{<x,\gamma><y,\mathbf{y}>\}}(0Ayx) \ \vee \ \tau_M^{\emptyset}(1xBx))$$

$$= \forall \mathbf{y}(\neg \tau_M^{\{<x,\gamma><y,\mathbf{y}>\}}(Ayx) \ \vee \ \exists \mathbf{x} \tau_M^{\{<x,\mathbf{x}>\}}(Bx))$$

$$= \forall \mathbf{y}(\neg \underline{A}(\tau_M^{\{<x,\gamma><y,\mathbf{y}>\}}(y), \tau_M^{\{<x,\gamma><y,\mathbf{y}>\}}(x)) \ \vee \ \exists \mathbf{x} \underline{B}(\tau_M^{\{<x,\mathbf{x}>\}}(x))$$

$$= \forall \mathbf{y}(\neg \underline{A}(\mathbf{y}, \gamma) \ \vee \ \exists \mathbf{x} \underline{B}(\mathbf{x})).$$

7 Towards Convention T

Theorem. Suppose f deals with all the untied tags of φ. Then for all R, $\mathcal{P}_M[\varphi, R, f] = R(1)$ is interdeducible with $\tau_M^f(\varphi)$.
Proof. By induction on the complexity of the play-chart φ.
Basis. Immediate from the basis clause in the definition of τ_M^f and the first boxed biconditional above.
Inductive Hypothesis. Suppose the result holds for all play-charts simpler than φ.
Inductive Step. Consider φ by cases. We shall deal here only with the case where φ is of the form 0ψ. We have

$$\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff not-}[\mathcal{P}_M[\psi, \overline{R}, f] = \overline{R}(1)].$$

By Inductive Hypothesis, we obtain

$$\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff not-}[\tau_M^f(\psi)].$$

By definition of τ_M^f, it follows that

$$\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff } \tau_M^f(0\psi).$$

The remaining cases are similar.

Corollary. If φ is a starting play-chart, then for all R, $\mathcal{P}_M[\varphi, R, \emptyset] = R(1)$ is interdeducible with $\tau^f_M(\varphi, \emptyset)$.

Note that $\tau^f_M(\varphi)$ does not involve R. It follows that if any $R : \{1, 2\} \xrightarrow{1\text{-}1} \{\alpha, \beta\}$ makes it the case that $\mathcal{P}[\varphi, R, f] = R(1)$, then every such R makes it so. Our theorem above therefore essentially correlates a relational property—call it $\mathcal{S}$—of φ, M and f with the condition $\tau^f_M(\varphi)$. And in the special case where φ is a starting play-chart (with all tags tied) we are correlating this relational property $\mathcal{S}$ of φ, M *and the null assignment* with the condition $\tau^f_M(\varphi, \emptyset)$. The latter is simply a statement about the model M. Let us further ignore the null assignment, and write $\mathcal{T}(\varphi, M)$ instead of $\mathcal{S}(\varphi, M, \emptyset)$; and write $\tau_M(\varphi)$ for $\tau^{\emptyset}_M(\varphi)$.

We have therefore fulfilled a model-relative form of Tarski's material adequacy condition on definitions of would-be truth-predicates:

> for every play-chart φ with no tags tied, and for every playing-field M, $\mathcal{T}(\varphi, M)$ is interdeducible with $\tau_M(\varphi)$.

Now Tarski took the view that if $\tau_{(M)}$ was a *translation mapping* (relative to M) from *what were regarded as* object-language sentences into (what would accordingly be regarded as) the *meta*language, then fulfillment of his material adequacy condition ensured that $\mathcal{T}(\varphi, M)$ would capture the truth of φ in M.

We have not yet, however, admitted that our 'play-charts' are sentences—that is, interpretable as statements about the 'playing-field' M. We have been treating them as non-semantic, albeit quasi-syntactic structures. Let us therefore take the leap, and allow play-charts to be *interpreted* in the obvious way given by τ above: basic play-charts are atomic predications; 0 is negation; 1 is disjunction; 2 is conjunction; $1x$ is the existential quantifier; and $2x$ is the universal quantifier. We realize that play-charts, as they were presented above, are in Polish notation. Tags are variables; the tied and untied ones are bound and free, respectively. Everything now falls into place. Courtesy of Tarski's adequacy condition on theories of truth, we can now regard the game-theoretical property

$$\mathcal{P}_M[\varphi, R, \emptyset] = R(1)$$

as that of *the truth of φ in M*. And that brings us now to the conjecture foreshadowed in the introduction.

8 A conjecture concerning truth and translation

The philosophical payoff may be more profound than is suggested by these remarks. It is commonly maintained that Tarski took the notion of translation (i.e., meaning-preserving mapping) as given, and used that notion to ensure that fulfillment of the adequacy condition entailed that the defined predicate was indeed a truth-predicate. For Tarski, therefore, the conceptual route would be from translation (meaning) to truth, courtesy of the T-schema.

It is also commonly maintained that Davidson 1967 reversed the direction of conceptual dependency. Davidson sought observational constraints on the postulation of (conjectural) biconditionals of the Tarskian form. In such postulated biconditionals, the predicate on the left *would already be interpreted as the truth-predicate.* Consequently, the empirically (albeit holistically) confirmed fulfillment of the adequacy condition would entail that the right-hand sides of the Tarskian biconditionals could be taken as the meanings of the object-language sentences referred to on the left-hand sides. For Davidson, therefore, the conceptual route would be from truth to meaning, courtesy of the T-schema. The standard wisdom, therefore, appears to be that if one takes either one of the notions of truth or translation (meaning-specification) as given, then Tarski's adequacy condition delivers the other notion.

A more radical suggestion arising from the foregoing treatment of games on *uninterpreted* quasi-syntactic structures is this: truth and translation (or: truth and meaning) are conceptually coeval. There is only one way of solving for each of them, simultaneously, and indeed uniquely, from fulfillment of the material adequacy condition. One does not need one of these notions to be given or fixed in advance in order then to determine the other notion. Rather, given the provability of all instances of

$$\mathcal{T}(\varphi, M) \leftrightarrow \tau_M(\varphi)$$

one can conclude—subject only to certain constraints on τ_M—*both* that $T(\quad, M)$ is the M-relative truth-predicate, *and* that τ_M is an M-relative *translation mapping* from (what can now obviously be regarded as *sentences* of an object-language) to (what we can now call) the metalanguage.

The constraints that need to be met by τ_M are not allowed to involve or presuppose the concepts of truth or translation (or meaning). But they will manifestly be met by the mapping we defined above (for example). Logicians assign a special (syntactic) meaning to the notion of a language. One important feature of a language is that it is expressively *closed.* That is, it contains the result of applying any of its expression-forming operators to any of its expressions of the appropriate syntactic classes. Thus, if a language contained the universal quantifier $\forall$ along with the negation operator $\neg$ and a monadic predicate F, it would have to contain all sentences that could be constructed by means of those expressions (plus variables): $\forall xFx$, $\forall x\neg Fx$, $\neg\forall xFx, \ldots$ and so on. A *sublanguage* of a language is any fragment thereof that is closed under the operation of certain of its expression-forming operators.

Now, one striking feature of the mapping τ_M defined above is that it mapped play-charts *onto* a sublanguage of the formalized language of games. Let us, by way of anticipation, call the latter the *metalanguage*, and let us also speak of sentences of the object language rather than speak of play-charts. The mapping need not be one-one; for two different predicates of the object language might map to one and the same predicate of the metalanguage. Nevertheless, the range of the mapping enjoys a certain richness or repleteness: everything that one 'might wish to say' using only grammatical and expressive resources of (metalinguistic) sentences already mapped onto can itself be mapped onto.[4] Note that τ_M is also 'M-relative' in that quantifiers in the sentences in its range are restricted to M, and atomic predications in its domain are mapped to relational claims about individuals within M. The condition of M-relativity can be spelled out purely syntactically.

Our sharpened version of the thesis of coeval conceptual determination is therefore as follows:

> Let τ_M be an M-relative homomorphic mapping from sentences of the object language onto a sublanguage of the metalanguage. Then the provability of all instances of
>
> $$\mathcal{T}(\varphi, M) \leftrightarrow \tau_M(\varphi)$$

[4]We must be careful, however, not to include any resources in the metalanguage that are embedded only in complex metalinguistic expressions that happen to be mapped onto by atomic formulas of the object language, but that do not occur in dominant position in expressions mapped onto by complex formulas of the object language.

suffices to determine that $T(\ ,M)$ is the M-relative truth-predicate and that $\tau_M(\varphi)$ is the M-relative translation of φ into the metalanguage.

9 Assessing Hintikka's criticisms

I have gone into greater detail here than I did in *Natural Logic* in order to bring out more clearly the subtlety of the philosophical insights to be gleaned from game theory's treatment of truth-conditions. The whole point of the game-theoretical treatment in *Natural Logic* was that the game-theoretical aspects were motivated *sui generis*, and a notation developed to keep elegant track of the barest materials involved in specifying a state of play in the game. Instead of Hintikka's own asymmetric conception of the players as Myself v. Nature (which conjures up images of the hapless individual pitted against merciless forces), I introduced the notion of a *role-assignment* R mapping the truth-values T and F one-one to the two players playing the game. This allowed a nice symmetrization (a levelling of the playing-field, as it were), and made better sense of players swapping roles. After swapping roles, the earlier role-assignment R changes to $\overline{R}$. One more role-swap, and we are back with R.

In *Natural Logic*, an inductive definition was provided of the notion

> the possessor of a winning strategy on formula $\varphi(x_1,\dots\ x_n)$ (in model M, relative to a role-assignment R, and relative to an assignment f of individuals in M to the free variables $x_1,\dots\ x_n$) is $R(T)$.

An equivalent but less ontologically committing way of expressing the same notion is:

> in the game on formula $\varphi(x_1,\dots\ x_n)$ (in model M, relative to a role-assignment R, and relative to an assignment f of individuals in M to the free variables $x_1,\dots\ x_n$), it is $R(T)$ who is in principle able to win (regardless of what his opponent does).

The clauses in that definition wrote themselves upon reflection on nothing but the internal, game-theoretic considerations. (In this paper, however, I have supplied the extra detail needed in order to reach those clauses as the ones boxed above, once one strips one's conceptual starting-point down to absolute semantical neutrality.) The reader of *Natural Logic* was in effect asked to stand back, suppress the double mention of R, suppress the occurrence of the constant T, and imagine the notion re-written as

$$\text{PREDICATE}(\varphi(x_1, \ldots\ x_n), M, f).$$

The resulting clauses (which had been motivated purely game-theoretically)[5] then struck one in the eye as none other than Tarski's clauses in his famous inductive definition of satisfaction. PREDICATE, it turned out, was none other than SATISFIES! This elegant 'morphing' of the game theoretical into the semantical showed directly how the concept of game-winnability (or: strategy-possession) had *semantic* content. The homology was so direct that it would have been laboring the obvious to 'prove' that the notions coincided. The proof would have been a rote exercise, once the required insight, just explained, had been attained.

On encountering Hintikka's unexpected complaint, however, that my truth-conditions have

> nothing to do with game-theoretical semantics except for the fact that Tennant uses the vocabulary of games in formulating them,

I thought it worth explaining in more detail how this is not so. I did not simply 'dress up' a standard truth-definition in game-theoretical vocabulary. Rather, one can motivate the notion of being in a winning position on an uninterpreted play-chart, and then discover, from what one is able to deduce from genuinely game-theoretic considerations, that, once the play-charts are interpreted as claims about M, their truth-conditions turn out to have been characterized game-theoretically.

Thus, in reply to Hintikka on this score, I hold it up as a virtue of the game-theoretical treatment given in *Natural Logic* that it so neatly delivered Tarskian truth and satisfaction, and did so without appeal to the axiom of choice. The notion of game-winnability finds perfect (and inductively definable) expression without any explicit ontologizing, as would be involved if one were to equate winnability with there being a winning strategy.

[5] So much for Hintikka's claim that my truth-conditions 'have nothing to do with game-theoretical semantics'!

One needed to carve the concepts at the right joints, as it were. The treatment in *Natural Logic* found just the right salient point of contact between game-theory and semantics. This is the equation of winnability with satisfaction. To have gone any further than that—in particular, to have quantified over strategies—would have been to introduce existential presuppositions that, arguably, had no legitimate place in the semantics of ordinary first-order logic. Occam's Razor is an important tool, even in the cut-throat world of philosophical logic.

Hintikka makes the mistake of thinking that anyone who does not do game-theoretical semantics in exactly his preferred way does not understand anything about it. I shall leave it to the reader to judge whether a lack of understanding on my part is evident from (i) the cleaner treatment of classical semantics offered in *Natural Logic*, (ii) the adaptation of that treatment to intuitionistic logic (via effective winning strategies) that was offered in Tennant 1979[6] and (iii) the many criticisms that I offered, in *Games*, of Hintikka's claims for IF-logic in his book *PMR*. The book contained many technical mistakes both in the exposition of IF-logic and in its various applications. It also made many exaggerated philosophical claims for the foundational significance of partially ordered quantifiers, semantically interpreted by means of games of imperfect information. *Games* took up matters of substantive detail, which Hintikka has not yet rebutted in print.

Hintikka writes that I am 'not imperceptive', but 'in denial'. I would not profess to be as well-versed in psychoanalysis as Hintikka. But I would offer the reminder, in conclusion, that one of the points made in *Games* (at pp. 107–110) was that Hintikka's IF-logic deprives us of a real negation operator with which to make denials. Finally, I would thank Hintikka for having written such colorful prose that at least a few more people might have their interest piqued enough to read *Games*.

[6] As noted on p. 114 of *Games*, Hintikka's adaptation of his own game-theoretical semantics for intuitionism does not guarantee player T a winning strategy on the intuitionistic logical truth $\neg\forall x(Px \wedge \neg Px)$!

References

Davidson, Donald. 1967. Truth and Meaning. *Synthese*, vol. 17, pp. 304–23.

Lorenzen, Paul and Lorenz, Kuno. 1978. *Dialogische Logik.* Wissensschaftliche Buchgesellschaft, Darmstadt.

Hintikka, Jaakko. 1973. *Logic, Language Games and Information.* Clarendon Press, Oxford.

Hintikka, Jaakko. 1996. *The Principles of Mathematics Revisited.* Cambridge University Press, Cambridge.

Hintikka, Jaakko. 2000. Game-Theoretical Semantics as a Challenge to Proof Theory. *Nordic Journal of Philosophical Logic*, vol. 4, no. 2, pp. 127–141.

Tennant, Neil. 1978. *Natural Logic.* Edinburgh University Press, Edinburgh.

Tennant, Neil. 1979. 'Language Games and Intuitionism'. *Synthese*, vol. 42, pp. 297–314.

Tennant, Neil. 1998. Games Some People Would Have All of Us Play. *Philosophia Mathematica*, series III, vol. 6, pp. 90–115.