Wojciech Załuski  
Chair for Philosophy of Law and Legal Ethics  
Department of Law, Jagiellonian University, Krakow

Game Theory and Legal Interpretation

**Abstract:** The goal of the paper is to present various ways in which game theory can be used to explicate the problems of legal interpretation. Three problems of legal interpretation are discussed: (1) what are the reasons for assuming that it is possible to distinguish on purely linguistic grounds between admissible and inadmissible meanings of a legal norm, i.e., for assuming that there are semantic limits to legal interpretation?; (2) what legal interpretation can be dubbed ‘proper’; in other words, how to choose one meaning of a legal norm from among a set of admissible meanings of this norm?; (3) why do judges often fail to choose a proper interpretation even though they know this interpretation? It is argued that game theory can be used to justify the claim that there exist semantic limits to legal interpretation, to explicate two main conceptions of a proper legal interpretation – intentionalism non-intentionalism, and to model the strategic behaviour of judges which consists in that they do not choose the proper interpretation even though they know it.

1. **Game theory in legal analysis**

Game theory can be applied in legal analysis in a descriptive and normative manner. In its descriptive usage game theory is treated as a tool for explaining and thereby predicting human behaviour, and in its normative usage – as a tool for determining the content of normative concepts, especially, the concept of justice. Given that game theory can be applied both in the area of legal philosophy and legal dogmatics, one can distinguish four ways of applying game theory in legal analysis: (a) normative in legal philosophy, (b) normative in legal dogmatics, (c) descriptive in legal dogmatics, (d) descriptive in legal philosophy. Arguably, as yet the applications of game theory in the areas (c) and (d) have proved to be most fruitful from among the possible application of game theory in legal analysis. As for (c): game theory has been extensively applied in legal dogmatics (in the context of various branches of law, e.g., tort law, contract law, competition law, bankruptcy law, consumer law) in order to predict how human being react to possible legal rules and thereby to choose the economically efficient legal rules. As for (d): game theory was also used in legal philosophy, e.g., in order to describe the function and emergence of law and to resolve or at least elucidate the problem of normativity. One of the well-known conclusions of this type of application of game theory is that from the standpoint of game theory law is a mechanism for solving collective action problems. Now, it seems that one of the important problems of legal

---

1 This area of application of game theory is covered in Baird, Gertner, Picker 2000.  
philosophy which has not been analyzed by means of game-theoretical tools is the problem of legal interpretation. The range of questions related to legal interpretation is very broad. In the following considerations I intend to focus mainly on some of them. Prior to formulating these questions, it is necessary to specify how the very concept of legal interpretation will be understood in the following considerations.

2. Legal interpretation

Legal interpretation can be defined as a process aimed at deciding which meaning of a given legal norm from among multiple *admissible* meanings of this norm distinguished on purely linguistic grounds is the most plausible (or the proper or the optimal) one. It is therefore a process of attributing the most plausible meaning to a legal norm. The effect of this process, i.e., the final choice of the most plausible meaning of a given legal norm, can be called ‘an interpretational decision’. It is clear that making an interpretation decision enables judges to decide whether a given legal case falls within the range of a given norm. It should be emphasized that the above understanding of legal interpretation assumes that, from a theoretical viewpoint, legal interpretation is a two-stage process: at the first stage a set of admissible meanings of a given legal norm is distinguished, and at the second stage the most plausible meaning from this set is selected. I have used the phrase ‘from a theoretical viewpoint’ because in concrete cases the judge does not have to go through both stages of the process: she may make only a *partial reconstruction* of the most plausible meaning of a legal norm if it is sufficient for deciding whether the case he is dealing with can be subsumed under a given norm. In the following sections I shall deal with the following three questions regarding legal interpretation:

(1) What are the reasons for assuming that it is possible to distinguish on purely linguistic grounds between admissible and inadmissible meanings of a legal norm?

(2) What, if any, legal interpretation can be dubbed ‘proper’ (or the most plausible or optimal); in other words, how to choose one meaning of a legal norm from among a set of admissible meanings of this norm?

(3) Why do judges often fail to choose the proper interpretation even though they know this interpretation?
The first question is about the existence of the semantic limits of a legal interpretation; the second question is about the existence of criteria for selecting the most plausible legal interpretation. The truthfulness of the thesis about the existence of semantic limits to legal interpretation is a presupposition of the truthfulness of the thesis about the existence of criteria for selecting the most plausible legal interpretation. Positive answers to these two questions can be viewed as substantiating the claim about the objectivity of legal interpretation (I shall introduce some further distinctions within the notion of the objectivity of legal interpretation in section 4). In the following sections I shall reflect on how game theory can be used in answering those questions or at least in explicating various answers given to these questions. In the remainder of the present section, I would like to make some additional remarks on the very process of legal interpretation.

The process of legal interpretation can be pursued in abstracto, i.e., as a theoretical exercise made in detachment from a concrete legal case (the results of such interpretation are not legally binding) or in concreto, i.e., either made in the context of concrete legal case but not by judges (so the results of such interpretation are not legally binding) or made in the context of concrete legal case by judges (so the results of such interpretation are legally binding). The analyses that follow can be referred to all those types of legal interpretation, though I shall treat as a paradigmatic type of legal interpretation the interpretation made by judges in concreto. Another question is whether each legal norm requires an interpretation. Many legal scholars hold the view that not all legal norms require an interpretation. This view is sometimes summarized by the dictum clara non sunt interpretanda. According to this view, only those legal norms whose meaning is ‘unclear’ require an interpretation. States this way, though, this view is deeply mistaken. There are no legal norms whose meaning is clear in the sense that one can be certain that they will never require an interpretation. For virtually each norm one can imagine a case with respect to which one may reasonably doubt as to whether the norm refers to the case or not. This is due to the fact that most linguistic terms are vague or at least potentially vague (e.g., exhibiting what F. Waismann called ‘open texture’, and H. L. A. Hart, following Waismann, called ‘openness’). However, there is some sense in the view saying that clara non sunt interpretanda, viz. in many concrete cases legal norms require no interpretation. Let me recall here Hart’s famous example of legal interpretation concerning a norm which bans the entrance of vehicles to the public park (see Hart 1994). The judge is faced with an interpretational problem whether bicycles are vehicles and thereby, by virtue of this norm, are banned from the public park. If we confine ourselves to a purely linguistic framework, it seems that we cannot exclude bicycle from the set of admissible
meanings of the norm banning vehicles from the public park. Even though bicycles do not have motor engine, they are means of transport and have wheels. Thus, there are some good, though not conclusive, reasons to regard a bicycle as an instance of a vehicle. However, to say this is not to forejudge that this is an apt interpretation of the norm. The problem whether it is an apt interpretation or not seems irresolvable on purely linguistic grounds. This kind of interpretational problem would not (typically) arise if the ‘candidate’ for a vehicle were not bicycle but, say, a truck (in Hart’s terminology: ‘a truck’ belongs to the ‘core of the determinate meaning’ of the word ‘vehicle’ and ‘a bicycle’ to its ‘penumbra of indeterminacy’). In summary, each norm requires a legal interpretation in the sense that, for each norm, one can imagine a case with respect to which it is uncertain whether this norm applies to it or not.

3. The semantic limits of legal interpretation

In order for legal interpretation to be objective there must exist semantic limits to it, i.e., it must be possible to distinguish on purely linguistic grounds a set of admissible meanings of a legal norm and this set must be non-trivial in the sense that there must exist also a set of inadmissible meaning of this norm. It can also be said that the claim about the existence of semantic limits to linguistic interpretation is equivalent to the claim that there exists semantic normativity, i.e., that there are correct and incorrect usages of linguistic expressions. The claim about the existence of semantic limits of legal interpretation can be defended on the common-sense ground, on the philosophical ground, and, arguably, also on the game-theoretic ground.

The common-sense argument for the existence of semantic limits to linguistic interpretation is the simple observation that in order to say anything meaningful it must be possible to say things incorrectly. In other words, our everyday communication implies the distinction between correct and incorrect usages of linguistic terms. Analogously, in the case of legal norms we are always capable of showing which interpretations of a given legal norm are obviously mistaken, i.e., not belonging to the set of admissible meanings of this norm. This argument from the ‘existence of semantic errors’ appears to strongly support the thesis about the existence of semantic limits to legal interpretation.

The philosophical argument for the existence of semantic limits to linguistic interpretation in fact a refinement of the common-sense argument: the refinement consists in providing a theory of meaning, i.e., specifying what it is in the nature of meaning that creates
semantic limits to linguistic interpretation. In the philosophical discussions about the ‘meaning of meaning’ various theories of meaning have been proposed. The theories can be divided, for our purposes, into those that reject the claim about the existence of semantic limits, i.e., those that in fact reject the very notion of meaning, and those that accept this claim. The following diagram presents a (simplifying) classification of these theories:

**Fig. 1. Theories of meaning**

Without going into the details of the controversies between advocates of those various theories, let me note that even though most meaning theorists are in strong disagreement what meaning exactly is, they usually agree (with the exception of postmodern theorists who assume the ‘deconstructivist’ theories of meaning) that meanings exist and that thereby there exist semantic limits to linguistic interpretation.

The core intuition of the game-theoretical analyses of meaning is that meanings were shaped within the society and that the final meanings of linguistic expression are the equilibria outcomes of recurrent coordination games. An interesting philosophical consequence of these analyses is that, given the way meanings emerged, one may justifiably expect that they are objective or, perhaps more aptly, intersubjective, i.e. similarly understood by the members of a society, but also to a certain degree subjective, i.e., unclear, open to manifold interpretations, as they arose in an evolutionary process on a trial and error basis and were not decreed by some rational entity. In other words, linguistic expressions are often not
clear but there exist semantic limits to their interpretation. Let me present here a simple game-theoretical model (a signaling game) of the emergence of meanings proposed by David Lewis (1986).³

In the game there are two players – the Sender and the Receiver. The sender has some private information about the world and wants to share it with the Receiver so that the Receiver could make a better decision. The game is one of pure coordination, i.e., there is no conflict of interests between the players: both the Sender and the Receiver want the Receiver’s decision to be optimal. The Sender can send various messages to the Receiver but none of these messages has a given meaning as yet. The number of states of the world, S, messages, M, and acts, A is equal. The game can be set out in the following matrix:

\[
\begin{array}{c|ccc}
\text{State} & \text{Act I} & \text{Act II} & \text{Act III} \\
\hline
\text{State 1} & 1, 1 & 0, 0 & 0, 0 \\
\text{State 2} & 0, 0 & 1, 1 & 0, 0 \\
\text{State 3} & 0, 0 & 0, 0 & 1, 1 \\
\end{array}
\]

*Fig. 2. A signaling game: the emergence of meanings*

The first payoff in each cell is the Sender’s payoff, the second payoff is the Receiver’s payoff. The matrix reflects the fact that the game is one of common interest. The strategies in this game are described by Brian Skyrms in the following way:

‘A Sender’s strategy in this game is a rule that associates each state with a message to be sent in that state; a Receiver’s strategy associates each message with an act to be taken if the message has been received. Sender’s strategy and Receiver’s strategy taken together associate an act taken by the Receiver with each state of the world. If, for every state, the act taken is optimal for that state, the combination of Sender’s strategy and Receiver’s strategy is called a signaling system (2004, pp. 52-53).’

Thus, signaling systems guarantee the optimal transmission of information. Skyrms gives the following examples of signaling systems for this game:

(a) Sender’s Strategy                      Receiver’s Strategy
    \[ S_1 \rightarrow M_1 \]                      \[ M_1 \rightarrow A_1 \]
    \[ S_2 \rightarrow M_2 \]                      \[ M_2 \rightarrow A_2 \]
    \[ S_3 \rightarrow M_3 \]                      \[ M_3 \rightarrow A_3 \]

³ My presentation is based also on Skyrms 2004, pp. 52-55.
Signaling systems are strict Nash equilibria of signaling games (i.e., a unilateral deviation from a signaling system by a given player not only does not make her better off but also makes her worse off) and thereby conventions in Lewis’s sense. They are also evolutionary stable strategies (where a strategy is defined as embracing both the Sender’s strategy and the Receiver’s strategy: each player is therefore seen as playing in both roles). It is to be stressed that signaling systems are the only strict Nash equilibria of this game. The other Nash equilibria of this game are weak, e.g.:4

(c) A totally non-communicative Nash equilibrium:

<table>
<thead>
<tr>
<th>Sender’s Strategy</th>
<th>Receiver’s Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1→ M3</td>
<td>M3→ A1</td>
</tr>
<tr>
<td>S2→ M1</td>
<td>M1→ A2</td>
</tr>
<tr>
<td>S3→ M2</td>
<td>M2→ A3</td>
</tr>
</tbody>
</table>

(d) A Nash equilibrium in which partial information is transmitted:

<table>
<thead>
<tr>
<th>Sender’s Strategy</th>
<th>Receiver’s Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1→ M1</td>
<td>M1→ A1</td>
</tr>
<tr>
<td>S2→ M1</td>
<td>M2→ A1</td>
</tr>
<tr>
<td>S3→ M3</td>
<td>M3→ A3</td>
</tr>
</tbody>
</table>

Since, as we have seen, one can create other equally good signaling systems by permuting the messages, the main problem of Lewis’s account is how to explain the choice of one of these systems. Skyrms’s answer to this problem is as follows:

‘In Lewis signaling games, which are games of common interest, evolutionary dynamics, learning dynamics, and almost any reasonable sort of adaptive dynamics lead to successful coordination on a signaling system equilibrium. In the absence of natural salience, which signaling system emerges depends on the vicissitudes of initial conditions and chance aspects of the process. But some signaling system does evolve because signaling systems are powerful attractors in the dynamics, and other Nash equilibria of the game are dynamically unstable (2004, p. 59).’

4 Cf. Skyrms 2004, p. 54.
In summary, the above account of the origins and the nature of meanings assumes that meanings emerge from social interactions and are conventions in Lewis’s sense. It is worth stressing that this account enables one not only to explain how linguistic conventions of meaning emerge but also how they can be maintained (they are maintained because they are equilibria of coordination games). As mentioned, this account seems to provide a strong case for the claim that meanings are stable and objective.

Let me now pass to the second question: about the criteria for selecting a proper meaning from among a set of admissible meanings.

4. The criteria for selecting a proper meaning

When a legal interpretation can be called ‘proper’, i.e., leading to the choice of the proper meaning? In other words, what are the conditions of a legitimacy of a legal interpretation besides its consistency with linguistic conventions? There might be no such criteria, i.e., that even though legal interpretation has semantic limits, there might be no good reasons to select meaning $M_i$ rather than meaning $M_j$ from the set of admissible meanings of a given legal norm. The thesis which assumes that there are semantic limits to legal interpretation but there are no criteria for selecting one – proper – meaning from among all admissible meanings can be called ‘the thesis about the weak objectivity of legal interpretation’. The thesis which assumes that there are no semantic limits to legal interpretation (and that thereby there are also no criteria for selecting one – proper – meaning from among all admissible meanings) can be called ‘the thesis about the subjectivity of legal interpretation’. The thesis which assumes that there are semantic limits to legal interpretation and there are good criteria for selecting one – proper – meaning from among all admissible meanings can be called ‘the thesis about the strong objectivity of legal interpretation’. In this section I shall present two conceptions which assume the thesis about the strong objectivity of legal interpretation; one view is called ‘intentionalism’, and the other one I shall call ‘non-intentionalism’. They differ in the criteria they specify for selecting a proper meaning from among a set of admissible meanings. Intentionalism says that a meaning of a norm $n$ is proper if and only if it is in accordance with the intentions of the legislator who has enacted this norm. Accordingly, in order to find a proper meaning of a given norm, one has to ascertain what meaning to a given norm was assigned by the legislator who enacted this norm. (One might raise an objection that this conception does not presuppose the claim about the existence of semantic limits to a legal interpretation, as it only says that a proper meaning of a
norm is the one accepted by the legislator who enacted this norm. This objection, though, does not seem to be apt. If there were no semantic limits to legal interpretation, then the interpreter could not even start the very process of searching for the proper interpretation, given the virtually infinite multitude of possible meanings she would have to take into account). Non-intentionalism says that an interpretation of a norm \( n \) is proper if and only if it is in accordance with, or fits, social reality. I shall reflect on how these two conceptions can be explicated by means of game-theoretical tools. The interpretational problem that I shall be appealing to is the one generated by the article 150 (of the Polish Criminal Code) which regulates euthanasia. The content of this article is as follows:

\[
\begin{align*}
\text{§ 1. } & \text{Kto zabija człowieka na jego żądanie i pod wpływem współczucia dla niego, podlega karze pozbawienia} \\
& \text{wolności od 3 miesięcy do 5 lat.} \\
\text{§ 2. } & \text{W wyjątkowych wypadkach może zastosować nadzwyczajne złagodzenie kary, a nawet odstąpić od jej} \\
& \text{wymierzenia.}
\end{align*}
\]

(§ 1. Whoever kills a human being upon his or her request and out of compassion for him or her shall be punished with imprisonment from 3 months to 5 years. § 2. In exceptional cases the court may mitigate a punishment or even renounce from inflicting a punishment).

This article may give rise to various interpretational problems but, for ease of exposition, I shall be focusing only on the problem generated by the term ‘out of compassion’. The problem can be stated as follows: compassion with what type of suffering? Let us assume, again for ease of exposition, that there are two types of suffering: physical and purely psychic (i.e., not being at the same type physical; physical suffering can be aptly regarded as always being at the same time psychic). Thus, the set of admissible meanings can be presented as follows:

\[
\begin{align*}
M_1 &= \{\text{physical suffering}\} \\
M_2 &= \{\text{purely psychic suffering}\} \\
M_3 &= \{\text{physical suffering, purely psychic suffering}\}.
\end{align*}
\]

It is assumed that the above meanings are the only meanings admissible on the grounds of the linguistic conventions governing the usage of the term ‘compassion’. Intentionalism and non-intentionalism imply that it is possible to choose one – the proper – meaning from among these three meanings.

4.1. Intentionalism

It seems that one can explicate intentionalism by means of the tools of rational choice theory in two different ways depending on whether one view the legislator as a strategic
player or not. In the former case, the proper tool is decision theory, in the latter case – game theory.

**Intentionalism and decision theory**

Within this framework one can say that according to intentionalism legal interpretation is a *quasi*-game between a judge and a legislator. It is not a game in the strict game-theoretic sense because the legislator does not make any strategic move in it: the legislator only enacts a given norm *in a given form* and, as we assume, assigns a specific meaning to this norm (one of M1-M3) but her decision to enact the norm *in a given form* is not determined by her expectations on how this norm *in this form* may be interpreted by the judge. However, it is not a purely decisional problem either because the legislator can be assigned different payoffs depending on what concrete move is made by the judge. The *quasi*-game is similar to a game of coordination because there is no conflict of interest between the players: the judge wants to choose the same meaning as the one that was chosen by the legislator; in other terms, I assume that the judge accepts the intentionalist conception of the proper meaning. The way in which the *quasi*-game is perceived by the judge is captured by the following matrix:

<table>
<thead>
<tr>
<th>Judge /Legislator</th>
<th>M1, p</th>
<th>M2, q</th>
<th>M3, 1 – p – q</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>1, 1</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
<tr>
<td>M2</td>
<td>0, 0</td>
<td>1, 1</td>
<td>0, 0</td>
</tr>
<tr>
<td>M3</td>
<td>0, 0</td>
<td>0, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

*Fig. 3. Intentionalism as a *quasi*-game: the judge’s perspective*

The probabilities sum to 1 because M1 and M2 are to be read as, respectively, ‘only physical suffering’, ‘only purely psychic suffering’. The three possibilities are therefore mutually exclusive. The probabilities are to be interpreted subjectively: they reflect the judge’s ignorance about what meaning has been intended by the legislator; of course, the judge has already chosen the meaning so that from her perspective it does not make sense to ascribe (non-trivial) probabilities to various meanings. The *quasi*-game is perceived in a different way by the legislator (it is assumed that the legislator intends the norm to be interpreted as M1):
<table>
<thead>
<tr>
<th>Judge /Legislator</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>1, 1</td>
</tr>
<tr>
<td>M2</td>
<td>0, 0</td>
</tr>
<tr>
<td>M3</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Fig. 4. Intentionalism as a quasi-game: the judge’s perspective

Clearly, after gathering evidence for ascribing most plausible probabilities to the three meanings, the judge will make a choice that maximizes her expected utility, i.e., in the above case, will simply choose the meaning that she deems to have the highest probability of having been chosen by the legislator. It should be noted, though, the judge may not have to make a complete interpretation of the norm but only a partial one, i.e., one that allows her to decide whether a given case she is deciding subsumes under the norm. Assume, e.g., that the case to be decided is killing a Y by an X out of compassion with Y’s purely psychic suffering. In such a case the judge does not have to decide which meaning from the above three meanings is the proper one; what she has to do in order to decide the case is to establish whether the sum of probabilities of M2 and M3 (which admit of euthanasia out of compassion for purely psychic suffering) is higher than the probability of M1. If the sum is higher, then she can decide that the case subsumes under the norm. In the context of this case she may leave the question of whether M2 or M3 is the proper interpretation undecided (however, given that she is likely to be obliged to make decisions in other cases concerning euthanasia, she cannot in practice avoid making the complete interpretational decisions).

**Intentionalism and game theory**

The above account of the interaction between the judge and the legislator is based on a serious simplification which seems to reduce it to triviality: it assumes that the legislator can formulate the norm only in one possible way. The truly strategic interaction between the legislator and the judge arises if we reject this assumption and take into account various possible ways of formulating a norm by the legislator. A model that captures this fact will be more realistic than the previous one. I present such a model below; the model is inspired by the model proposed by P. Parikh (2001, p. 30-31) in an extra-legal contexts.
Fig. 5. Intentionalism as a game between the judge and the legislator.

The legislator has to make a choice between various formulations of the norm concerning euthanasia:

\( m \): Whoever kills a human being upon his or her request and out of compassion with his or her physical suffering (but not out of compassion with his or her purely psychic suffering) shall be punished with imprisonment from 3 months to 5 years.

\( n \): Whoever kills a human being upon his or her request and out of compassion with him or her shall be punished with imprisonment from 3 months to 5 years.

\( o \): Whoever kills a human being upon his or her request and out of compassion with his or her purely psychic suffering (but not out of compassion with his or her purely physical suffering) shall be punished with imprisonment from 3 months to 5 years.
p: Whoever kills a human being upon his or her request and out of compassion with his or her physical suffering or out of compassion with his or her purely psychic suffering shall be punished with imprisonment from 3 months to 5 years.

The formulations m, o, p are non-vague but very awkward, whereas the formulation n (assumed by the Polish legislator) is vague but elegant. The legislator may prefer M1, M2, or M3. It is assumed that if the legislator chooses the formulations m, o, or p, the judge knows the meaning accepted by the legislator and in the game branches generated by the choices of these formulations the game is one with perfect information. However, if the legislator chooses the vague formulation of the norm, the game will be one with imperfect information: the information set S2 of the judge contains three elements and thereby the judge, when making her decision, is ignorance of the meaning assigned to the norm by the legislator. The result of the game will depend on utility functions of the players and the probabilities of preferring M1, M2, and M2 by the legislator who has selected the vague formulation n. By way of example, let us assume that the payoffs for the judge and the legislator is 20 when they choose the same meaning under the vague formulation n of the norm regarding euthanasia, and 8 under any of the non-vague formulations of the norm; I therefore assume that the costs of enacting a non-vague but awkwardly formulated norm amount to 12); I also assume that the probabilities of preferring M1, M2, M3 by the legislator enacting a vague norm are, respectively, 1/2, 1/4, 1/4 (of course, one of the main problems with constructing a realistic model of analogous games is to justify, not just assume, the payoffs and probability values introduced into the model). It is clear that with these values the Nash equilibrium of this game is the following combination of strategies:

Legislator: if M1, then n; if M2, then o; if M3, then p.

Judge: if in S1, then M1; if in S2, then M1; if in S3, then M2; if in S4, then M3.

This combination does not create for any player an incentive to unilaterally change her strategy. For instance, if the legislator prefers M1 and chooses n, she can expect to receive the payoff 10 (1/2 · 20) which is more than 8 she would receive if she chose an alternative formulation m; and if the judge is in S2 (as a result of the legislator’s choice of n), she will choose M1 because her expected utility from this choice is 10 (rather than 5 – as it is for her choice in S2 of M2 or M3); thus ‘if M1, then n’ and ‘if in S2, then M1’ are best responses to each other; an analogous argumentation for the other elements of the above combination shows that the entire combination forms a Nash equilibrium.
The above analysis was aimed at revealing the strategic structure of the interaction between the legislator and the judge on the assumption that the judge’s purpose is to find a proper interpretational decision in accordance with the criteria provided by intentionalism. I have not tackled the philosophical problem of evaluating intentionalism as a theory of a proper legal interpretation, as this problem lies beyond the scope of this paper. In this place I would like to limit myself to enumerating only two of numerous difficulties encountered by this theory. First, it is by no means clear whether it is justifiable to speak about the intentions of the legislator, as the legislator is usually a ‘group entity’. Second, more fundamentally, it is not easy to find good reasons for which the judges should be guided by the legislators’ intentions. Finishing this section, I shall present an opinion of Andrei Marmor’s, which – in my view – is a very convincing answer to the second question:

‘The point I wanted to make is strictly conditional: if, and only if, a certain law is justified on the basis of the expertise branch of the normal justification thesis (i.e. only if the legislator can be regarded as possessing special – expert – knowledge about the regulated issues – the knowledge inaccessible to the judge – W.Z), would it make sense to defer to the legislature’s intentions in the interpretation of the law, that is, to the extent that there is, in fact, such an intention and it can clarify something that needs clarification. It is not part of my argument to insist that this is likely to happen very often (Marmor, 2005, p. 139)’.

4.2. Non-intentionalism

As mentioned before, non-intentionalism says that an interpretation of a norm n is proper if and only if it is in accordance with, or fits, social reality. This vague definition can be made clearer by explicating the notion of social reality. By social reality I shall mean a set of social rules; I treat them as constituting social reality (to the extent that is relevant for legal interpretation). Traditionally, four types of such rules are distinguished: linguistic, legal-systemic, moral and economic. Linguistic rules define the set of admissible meanings of a legal norm. Legal-systemic rules require that the selected meaning be consistent with the legal system to which an interpreted norm belongs. The selected meaning of an interpreted norm should also be consistent with moral rules that hold in a given society and should not violate the demands of economic efficiency. Now, the meaning of an interpreted legal norm that is consistent with these four requirements fits the social reality and thereby is the proper (or the most plausible) legal interpretation.

How can this conception of a proper legal interpretation be explicated or at least presented by means of game-theoretical tools? I would like to suggest that legal interpretation can be viewed as a quasi-game between an interpreter (judge) and social reality. The concept of a quasi-game I use to describe the interpreter’s activity is inspired by Hintikka’s semantic games, i.e., ‘games of nature’ intended to model an ‘interaction’ between ‘Myself’ (Verifier,
Scientist) and ‘Nature’ (Falsifier). To make my point clear, I must briefly present the main ideas of Hintikka’s game-theoretical semantics (GTS).\(^5\) GTS is semantics for quantificational (first-order) interpreted languages; it can be viewed as providing a link between language and reality. A language \(L\) is interpreted if there is a domain of individuals \(D\) on which all the predicates of the language \(L\) in question are interpreted. Thus, each atomic sentence constructed out of the predicates of \(L\) and of the proper names of members of \(D\) (it is assumed that proper names of members of \(D\) are the only singular terms of \(L\)) have a definite truth-value – true or false. Now, the task of GTS is to extend the truth value from these atomic sentences to all the sentences of \(L\). GTS realizes this task by introducing the concept of \(G(S)\) games, i.e., games correlated with sentence \(S\). The games \(G(S)\) can be viewed as an idealized process of verification in which one of the two players – ‘Myself’ (Verifier, Scientist) – wants to show that a sentence \(S\) is true and the other player – ‘Nature’ (Falsifier, Cartesian malin genie) – wants to show that it is false. ‘Myself’ wins if the game ends with a true atomic sentence and loses if the game ends with a false atomic sentence. Hintikka defines the truth independently of the course of any particular play of semantic games: a sentence \(S\) is true if there is a winning strategy for ‘Myself’ in a game \(G(S)\) correlated with sentence \(S\). In this definition the notion of strategy is understood in a strict game-theoretical sense. A player has a winning strategy if she wins the game no matter what her opponent does. The definition of truth provided by GTS is equivalent to the traditional definition of truth; as Hintikka writes:

‘It is most easily seen that if \(S\) is indeed true in the traditional sense, I can make my moves so that all the sentences \(S\) produced during the game are (apart from switches or roles induces by (G.¬) true in the traditional sense. Since this includes the outcome, I have a winning strategy. Conversely, if I have a winning strategy in \(G(S)\), it is easily seen that \(S\) is true in the traditional sense (Hintikka 1979b, p. 36).’\(^5\)

Hintikka also says that semantic games can be regarded as games of seeking and finding: ‘Nature’ is hostile and does not want to present the object ‘Myself’ wants. In order to find this object, ‘Myself’ must have a winning strategy. This name is especially apt if we take into account existential statements in the case of which the task of ‘Myself’ is to search for and find a suitable individual. Within the framework of GTS, the statement ‘\(p\) and \(q\)’ is treated as a universally quantified statement expression ‘Every one of the sentences \(p, q\) holds’. It is ‘Nature’ that chooses one of the sentences; thus, when playing the game \(G(p \text{ and } q)\), ‘Nature’ chooses whether the game should proceed as \(G(p)\) or \(G(q)\). A game of conjunction is won by the ‘Myself’ if both atomic sentences are true. A game of disjunction \(G(p \text{ or } q)\) proceeds in the following way: ‘Myself’ ‘decides whether the game should proceed

---

G(p) or G(q). The game is won by ‘Myself’ if at least one component of the disjunction is true. The game with a universally quantified statement \( G(\forall x(f)x) \) proceeds thus: ‘Nature’ chooses an object and provides a name \( a \) for it, and the game proceeds as \( G(f(a)) \). In the case of an existential quantified sentence \( G(\exists x(f)x) \) it is ‘Myself’ who chooses. In the game of negation \( G(\neg p) \), the roles of the players as defined by the above rules – are reversed and the game continues for \( p \).

Now, it seems that with terminology of GTS one can describe the basic idea of non-intentionalism in the following way: the judge proposes various meanings as ‘candidates’ for the proper interpretations of a legal norm, and the purpose of social reality is to falsify the claim of the interpreter that the meaning proposed by the judge fits social reality, i.e., is in consistent with: linguistic rules, with the systemic rules in a legal system, with the basic moral rules, and with the basic requirements of economic efficiency. The sentence \( S \) proposed by the judge whenever she proposes some meaning is: \( (L \land S \land M \land E) \), i.e., the judge claims that the meaning \( M \) she has selected is proper, i.e., consistent with the four types of rules mentioned above (and denoted, respectively, as \( L, S, M, E \)), and ‘social reality’ can choose any conjunct to falsify the interpreter’s claim. The basic idea of the above account of interpretation is presented in the following diagram:

![Diagram of non-intentionalism as a quasi-game of the judge with 'social reality'](imageURL)

*Fig. 6. Non-intentionalism as a quasi-game of the judge with 'social reality'*

If one takes into account the four requirements of a proper legal interpretation – linguistic (L), systemic (S), moral (M) and economic (E), and assumes that each requirement can be fulfilled or not (rather than fulfilled to a certain extent), one will easily notice that a proposed meaning may fail to fit social reality in 15 different ways:
Fig. 7. Ways in which proposed meanings may fail to fit ‘social reality’

The meanings 9-16 are not admissible, as they violate linguistic conventions. The judge who chooses such meanings does not interpret the law but creates it. From among the remaining meanings, only meaning 1 is proper: it satisfies all the requirements (linguistic, systemic, moral, and economic) and Pareto-dominates all the other meanings. The goal of the judge is to find this meaning.

The above proposal of explicating non-intentionalism can be criticized on various grounds. First, the application of Hintikka’s semantic games in the analysis of the process of legal interpretation may be criticized for its being based on superficial analogies between games of nature and the games played by the judges or as being non-illuminating, i.e., not saying anything new and interesting about the process of finding a proper meaning but only rephrasing the well-known facts. Second, and perhaps more importantly, the above proposal relies on a grossly simplifying assumption that the criteria for evaluating various possible meanings operate in an all-or-nothing fashion, i.e., that a given meaning may satisfy a given criterion or not rather than satisfy it to a higher or smaller degree. In point of fact, only linguistic criteria seem to operate in an all-or-nothing fashion, i.e., they may be fulfilled or not; as was mentioned before, the fulfillment of these criteria is regarded as a condition of an admissibility of a proposed meaning. Third, the above proposal assumes that the three criteria – legal-systemic, moral, and economic have equal weights, i.e., are equally important. The

<table>
<thead>
<tr>
<th>Meaning</th>
<th>L</th>
<th>S</th>
<th>M</th>
<th>E</th>
<th>Fit/Lack of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>LF</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>LF</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>LF</td>
</tr>
</tbody>
</table>
problem is that depending on what legal-philosophical views one assumes one may assign different weights to those criteria. Thus, in order to find a proper meaning one must apply these criteria in accordance with some decisional formula that assigns weights to those criteria. The general form of this decision formula is: \( \max [w_S(S), w_M(M), w_E(E)] \), where \( w_S \), \( w_M \), \( w_E \) are weights assigned to the various criteria by a given judge depending on the legal-philosophical views she assumes; the formula therefore prescribes making a balanced maximization of different values (legal-systemic, moral, and economic) while choosing a proper meaning of a norm.\(^6\) An adherent of legal positivism will assign the highest weight to \( S \), an adherent of the natural law views – to \( M \), and an adherent of *Law & Economics* to \( E \). The formula can be applied to those meanings that satisfy the linguistic criteria (which is a necessary condition of the plausibility of an interpretational decision). This ‘decision-theoretical’ account of non-intentionalism seems to be more plausible than the game-theoretical one which draws on the analogy between Hintikka’s semantic games and the process of legal interpretation.

5. **The strategic reasons for failing to select a proper meaning**

In the final section of this paper I would like to devote a few remarks to the problem of the reasons which judges may fail to make a proper legal interpretation. Of course, there are many such reasons. The most obvious one is that they may simply not know which interpretation is proper. The other one, which I shall analyze below, is that even though they know the proper interpretation they may be unwilling to make it, i.e., their reasons for not selecting a proper interpretation are strategic.

At the start let me make a distinction between individual and collective interpretational situation. An interpretational situation is collective if at least two legal actors are responsible for making an interpretational decision (though, of course, the number of decision-makers is usually odd). Collective interpretational situations may be strategic or not. Consider, for instance, a situation in which three judges have to decide whether a given behaviour of a person \( X \) constitutes a violation of a norm \( n \). The set of actions available to each judge is the same: each may say that \( X \) violated \( n \) or that \( X \) did not violate \( n \). If their choices are not unanimous, then the final decision is made by some mechanism of collective decision-making, e.g., the majority voting. However, the strategic element may appear here in the form

\(^6\) A sophisticated framework for making balanced maximization is provided by Giovanni Sartor in Sartor 2010.
of the so-called strategic voting, i.e., voting not in accordance with one’s preference-ordering over available options with a view to obtaining the best possible option given the preference-orderings of the other decision-makers. In my further considerations I will not be concerned with this type of interpretational situations. An interpretational situation is individual if only one legal actor is responsible for making an interpretational decision. The strategic behaviour may appear in this situation in at least two contexts. First, a judge may be concerned with what interpretational decision other judge – faced with a similar case - will make. The judges may be, for some reason, willing to make the same interpretational decision; e.g., they may derive especially utility from making the same decision, as deciding in a similar way lowers the probability of their decisions’ being reversed by the appellate court. Second, a judge’s decision may be determined by her expectations about what the appellate court would do in case the judge’s decision has been appealed against. The following two games illustrate these two cases.

(1) Two judges want to make the same decision regarding an interpretation of some legal norm. However, their preference-ordering of the possible meanings is different: each prefers a different meaning. Thus, they play the game known as the Battle of Sexes.

<table>
<thead>
<tr>
<th>Judge 1/Judge 2</th>
<th>M1</th>
<th>M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>2, 1</td>
<td>0, 0</td>
</tr>
<tr>
<td>M2</td>
<td>0, 0</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

*Fig. 8. The Battle of Sexes game as judges’ game*

If they can communicate, they can agree upon a probability mix \{(M1, M1), probability = 1/2; 1/2 (M2, M2), probability = 1/2\}. However if they cannot communicate, they may reach the worst outcomes (M2, M1) or (M1, M2).

(2) A judge must decide between Meaning 1 and Meaning 2. She prefers the former to the latter but knows that the appellate court’s preference-ordering is different. Thus, if the judge chooses Meaning 1, her decision will be overruled by the appellate court. The preference ordering of the judge is as follows: Supported Meaning 1 > Meaning 2 > Overruled Meaning 1 (and therefore Meaning 2 after overruling Meaning 1). The appellate court’s preference ordering is as follows: Meaning 2 = Overruled Meaning 1 (and thereby Meaning 2) > Supported Meaning 1 (it is therefore assumed that the appellate court wants only that Meaning 2 rather than Meaning 1 is reached independently of how it is reached – by or without overruling).
The rollback equilibrium (and thereby subgame perfect Nash equilibrium) of this game is \{Meaning 2, Overrule\}. Thus, even though the judge prefers Meaning 1 to Meaning 2, strategic considerations will lead her to choose Meaning 2. The equilibrium outcome of the game will be (1, 1). Thus, the judge will not choose the meaning she deems the best one because she does not want her decision to be overruled by the appellate court. The above model is intentionally very simple: it can complicated \textit{ad libitum}.

The above analysis leads to the conclusion that there are two main models of making interpretational decisions. The first one, which can be called ‘non-opportunistic’, says that the judge makes her interpretational decisions only in one stage – the stage of games or \textit{quasi}-games aimed at establishing the proper legal interpretation. The second one, which can be called ‘opportunistic’, says that the judge makes her interpretational decisions in two stages – the stage of games or \textit{quasi}-games aimed at establishing the proper legal interpretation, and the strategic stage at which the judge decides, taking into account the expected behaviour of other judges, whether to choose the proper legal interpretation or some other interpretation (which is not proper according to the ‘intentionalist’ or ‘non-intentionalist’ theory of proper legal interpretation). Accordingly, one can distinguish two types of judges: non-opportunistic, i.e., making their interpretational decisions at one stage, and opportunistic, i.e., making their interpretational decisions at two stages. In real life most judges makes their interpretational decisions sometimes in an opportunistic way, sometimes in a non-opportunistic way; there are few judges who are always opportunistic or always non-opportunistic.

Fig. 9. The game of the judge with the appellate court

---

7 More sophisticated models can be found, e.g., in Eskridge 1991a, Eskridge 1991b, Vanberg 2005).
References

Baird, C. and R. H. Gertner, R. C. Picker  

Eskridge, W.N.  

Hart, H.L.A.  

Hintikka, J.  

Hintikka, J., Symons. J  

Kornhauser, L.A.  

Lewis, D.  

Marmor, A.  

Parikh, P.  

Postema, G. J.  

Sartor, G.  

Skyrms, B.  

Ullmann-Margalit, Edna  