Working Papers
Institute of Mathematical Economics

Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

No. 92
John C. Harsanyi and Reinhard Selten
A Non-Cooperative Solution Theory with
Cooperative Applications
Chapter 2
Consequences of Desirable Properties
March 1980
Chapter 2. Consequences of Desirable Properties

The nature of the problem of equilibrium point selection in non-cooperative games does not seem to permit a satisfactory solution concept which can be characterized by a set of simple axioms. Nevertheless, it is useful to look at desirable properties which one might want to require and to explore their consequences.

Even if full scale axiomization cannot be achieved, important conclusions can be drawn from axiomatic considerations of limited scope. The simplest class of games where the equilibrium point selection problem occurs is that of all 2x2-games with two strong pure strategy equilibrium points; here the word "strong" is to be understood in the sense that a player loses by a deviation from his equilibrium strategy if the other players stick to their equilibrium strategies. A central notion of our theory, namely that of risk dominance can be fully axiomatized for this admittedly very restricted class of games.

It is also important to see that certain properties which may seem to be desirable at first glance cannot be achieved. As we shall see, it is impossible to define a continuous solution. Another impossibility result to be derived below concerns a way of subdividing one information set into two which we call "sequential agent splitting". An agent who has to choose between a, b, c is subdivided into two agents, one who first chooses between a and bc and another who then, if necessary, decides between b and c. Unfortunately, it is not possible to require that this kind of agent splitting should not essentially change the solution of the game without violating other axioms which we judge to be intuitively more compelling.

We shall also look at substructures of agent normal forms which are closely connected to subgames in the extensive form. These substructures, called cells, give rise to powerful requirements which reduce the task of finding a solution for general games to the task simpler one of finding a solution
for games without such substructures.

1. Continuity

Consider the class of all 2x1-games of the form shown in figure II-1. For $t \neq 0$ the game has only one equilibrium point, namely

A
\[
\begin{array}{c|c}
    t & 0 \\
    \hline
    0 & 1 \\
\end{array}
\]

B

Figure II-1: A class of 2x1-games

A for $t > 0$ and B for $t < 0$.

For $t = 0$ every mixed strategy of player 1 is an equilibrium strategy. Clearly, no solution concept can assign a unique equilibrium point to every game in the class in a continuous way. Not only player 1's strategy but also player 2's payoff must behave discontinuously as a function of $t$ at $t = 0$.

If a payoff parameter is varied continuously, some equilibrium points may suddenly disappear and others which have not been there before may suddenly appear. In order to show how this problem may arise in a less trivial way we add a further example. Consider the class of games given by figure II-2. Here for $t < -1$ the strategy combination Aa is the only equilibrium point

A
\[
\begin{array}{c|cc}
    & a & b \\
    \hline
    F & 2 & 0 \\
    G & 1-t & 0 \\
    H & 0 & 1+t \\
    I & 0 & 2 \\
\end{array}
\]

B

Figure II-2: A class of 2x2
of the game. For $-1 \leq t \leq +1$ both $A$ and $B$ are equilibrium points. Moreover, for $-1 < t < +1$ the game has a third equilibrium point in mixed strategies where player 1 uses $A$ with probability $2/(3+t)$ and player 2 uses $a$ with probability $(1+t)/(3+t)$. For $-1 < t < +1$ the game has no further equilibrium point. For $t = -1$ and $t = +1$ there are infinitely many equilibrium points, but this does not matter as far as our argument is concerned. Any function which assigns a unique equilibrium point to every game in the class must behave discontinuously with respect to $t$ at some point in the interval $-1 \leq t \leq +1$. This example differs from the preceding one inasmuch as for every $t$ the game has only a finite number of equilibrium points.

The game has no further equilibrium point. Any function which assigns a unique equilibrium point to every game in the class must behave discontinuously with respect to $t$ at some point in the interval $-1 \leq t \leq +1$. This example differs from the preceding one inasmuch as for every $t$ the game has only a finite number of equilibrium points.

It is now clear that a certain amount of discontinuity cannot be avoided in a theory of equilibrium point selection. Continuity considerations seem to be of little relevance for the problem.

2. Definitions and notations

Before we can go on to investigate further desirable properties of a non-cooperative solution theory we must introduce some definitions and notations.

Normal forms: A game in normal form $G = (\Phi, H)$ consists of a set of pure strategy continuations

\[
\Phi = \bigotimes_{i \in \mathbb{N}} \Phi_i
\]

and a payoff function $H$ which assigns a payoff vector

\[
H(\phi) = \left( H_i(\phi) \right)_i^{\infty}
\]

to every pure strategy combination $\phi = (\phi_i)_{i \in \mathbb{N}} \in \Phi$.

The lower index $\mathbb{N}$ indicates that $\phi$ contains one $\phi_i$ for every $i \in \mathbb{N}$ and $H(\phi)$ contains one $H_i(\phi)$ for every $i \in \mathbb{N}$. Player $i$'s payoff $H_i(\phi)$ for $\phi$ is a real number. The elements $\phi_i$ of $\phi_i$ are player $i$'s pure strategies. The sets $\Phi_i$ of pure strategies are finite.

In many cases the player set $\mathbb{N}$ will simply be set $\{1, \ldots, n\}$ of
the first \( n \) integers but since we must look at substructures of games which are games with fewer players it is convenient to define a normal form in such a way that the player set can be any non-empty finite subset of the set of positive integers. For \( N = \{1, \ldots, n\} \) we can write \( \phi = (\phi_1, \ldots, \phi_n) \) and \( H(\phi) = (H_1(\phi), \ldots, H_n(\phi)) \).

Often a game in normal form will simply be called a game where this can be done without risk of confusion. We shall mainly be concerned with such games even if extensive forms will be looked at occasionally in order to clarify conceptually important points. It must be kept in mind that in the framework of our theory a normal form must be interpreted as a perturbed agent normal form of an extensive game.

Further definitions in this section will refer to a fixed game \( G = (\phi, H) \).

Mixed strategies: A mixed strategy \( q_i \) of player \( i \) is a probability distribution over player \( i \)'s set \( \Phi_i \) of pure strategies. \( q_i(\phi_i) \) denotes the probability assigned to \( \phi_i \). No distinction is made between a pure strategy \( \phi_i \) and that mixed strategy which assigns probability 1 to \( \phi_i \) and 0 to all other pure strategies. The set of all mixed strategies \( q_i \) of player \( i \) is denoted by \( Q_i \).

A combination \( q = (q_i)_N \) of mixed strategies contains a mixed strategy \( q_i \) for every \( i \in N \). The set of all combinations \( q \) of this kind is denoted by \( Q \). For \( q = (q_i)_N \in Q \) and \( \phi = (\phi_i)_N \in \phi \), it is convenient to introduce the notation

\[
(2.3) \quad q(\phi) = \prod_{i \in N} q_i(\phi_i)
\]

In other words, \( q(\phi) \) is the product of all \( q_j(\phi_j) \) with \( j \in N \). The product \( q(\phi) \) is called the realization probability of \( \phi \) under \( q \). The definition of the payoff function \( H \) is extended from \( \phi \) to \( Q \) in the usual way:

\[
(2.4) \quad H(q) = \sum_{\phi \in \phi} q(\phi)H(\phi)
\]

It will be necessary to look at combinations of the form \( q_{-i} = (q_j)_{N-\{i\}} \), which contain one strategy for every player with the exception of \( i \). The index \(-i\) is used in order to designate such combinations which are called \( i \)-incomplete.
\( \mathcal{Q}_i \) denotes the set of all \( i \)-incomplete pure combinations and the symbol \( \mathcal{Q}_-i \) is used for the set of all \( i \)-incomplete mixed combinations. We use the notation \( q_i q_-i \) in order to describe that \( q \) which contains \( q_i \) and the components of \( q_-i \). If for all players with the exception of player \( i \) the strategies in \( q_-i \) agree with that in \( q \) we call \( q_-i \) the \( i \)-incomplete combination derived from \( q \).

**Best replies and equilibrium points:** \( r_i \in \mathcal{Q}_i \) is called a best reply to \( q_-i \in \mathcal{Q}_-i \) if we have

\[
H_i(r_i q_-i) = \max_{q_i \in \mathcal{Q}_i} H_i(q_i q_-i)
\]

(2.5)

It is a well-known fact of game theory that \( r_i \) is a best reply to \( q_-i \) if and only if every pure strategy \( \varphi_i \) with \( r_i(\varphi_i) > 0 \) is a best reply to \( q_-i \).

We say that \( r_i \) is a best reply to \( q \in \mathcal{Q} \) if \( r_i \) is a best reply to the \( i \)-incomplete combination \( q_-i \) derived from \( q \). A combination \( r \in \mathcal{Q} \) is called a vector best reply or shortly a best reply to \( q \in \mathcal{Q} \) if every \( r_i \) in \( r \) is a best reply to \( q \).

\( r_i \) is called a strong best reply to \( q_-i \) if \( r_i \) is the only best reply to \( q_-i \). In view of what has been said above, a strong best reply is always a pure strategy.

An equilibrium point is a strategy combination \( r \in \mathcal{Q} \) which is a best reply to itself.

An equilibrium point \( r \) is called strong for \( i \) if player \( i \)'s strategy \( r_i \) in \( r \) is a strong best reply to \( r \). A strong equilibrium point is an equilibrium point which is strong for every \( i \in \mathbb{N} \).

**Solution function:** Let \( \mathcal{G} \) be a class of games in normal form. A solution function \( L \) for \( \mathcal{G} \) is a function which assigns one of its equilibrium points \( r = L(G) \) to every \( G \in \mathcal{G} \).

In the following we shall look at desirable properties of solution functions and of concepts which are used in the definition of solution functions. We have already discussed
continuity in the previous section.

3. Invariance with respect to positive linear payoff transformation

The payoff of the players are von-Neumann-Morgenstern utilities. Interpersonal comparisons may be possible but they should not be considered as relevant for a non-cooperative solution theory where each player is assumed to be motivated by his own payoff exclusively.

Interpersonal utility comparisons are important for ethical theory but they have no room in a solution concept which is exclusively based on individualistic rationality assumptions.

Since von-Neumann-Morgenstern utilities are determined only up to positive linear transformations and since interpersonal comparisons are considered irrelevant, a game remains essentially unchanged if each player's payoff is subjected to a different positive linear transformation. This leads to the following definition of equivalence between games.

**Equivalence:** Two games \( G = (\phi, H) \) and \( G' = (\phi, H') \) with the same set \( \phi = \bigcup_{i \in \mathbb{N}} \phi_i \) of pure strategy combinations are equivalent if constants \( \alpha_i > 0 \) and \( \beta_i \) can be found for every \( i \in \mathbb{N} \) such that

\[
(2.6) \quad H_i' (\phi) = \alpha_i H_i (\phi) + \beta_i
\]

holds for every \( \phi \in \phi \) and every \( i \in \mathbb{N} \).

**Invariance with respect to positive linear payoff transformation:** A solution function \( L \) for a class \( \mathcal{G} \) of normal form games is invariant with respect to positive linear payoff transformations if for two equivalent games \( G \) and \( G' \) in \( \mathcal{G} \) we always have \( L(G) = L(G') \).

Invariance with respect to positive linear payoff transformations is a very important requirement. It is more than a
4. Symmetry

A rational theory of equilibrium point selection must determine a solution which is independent of strategically irrelevant features of the game. Names and numbers used to distinguish players and strategies should not matter \(^1\). Games which do not differ in other ways must be considered as isomorphic and should not be treated differently.

Invariance with respect to renaming of players and strategies may be looked upon as a symmetry property since its most important implication can be seen in the fact that the solution must reflect the symmetries of the game.

A renaming of players and strategies in a game \( G = (\phi, H) \) may be thought of as a system of one-to-one mappings \( f = (f_i)_{i \in N} \) where \( f_i \) maps player \( i \)'s pure strategy set \( \phi_i \) onto a new strategy set \( \psi_{\sigma(i)} \). Here \( \sigma(i) \) is a one-to-one mapping from the player set \( N \) onto a new player set \( N' \) and \( \sigma(i) \) is the new name of player \( i \). In this way a new game \( G' = (\psi, H') \) with \( \psi = \bigcup_{i \in N'} \psi_i \) arises.

We may look at \( f \) as a mapping from \( \phi \) to \( \psi \). This suggests the notation \( f(\phi) \) for that combination \( \phi \in \psi \) whose components are related by \( \phi_{\sigma(i)} = f_i(\psi_i) \) to those of \( \psi \). The new payoff function \( H' \) satisfies the condition

\[
(2.7) \quad H'_{\sigma(i)}(f(\phi)) = H_i(\phi)
\]

for every \( i \in N \) and every \( \phi \in \phi \).

An example is given in figure II-3.

It is convenient to adopt a notion of isomorphism which permits us to say that equivalent games are isomorphic. Therefore our definition of an isomorphism will involve a combination of a renaming with a system of positive linear payoff transformations.
Figure II-3: A renaming of players and strategies. The renaming may be thought of as performed in two steps. The first one is a renaming of player 2’s strategies which corresponds to an exchange of columns in the bimatrix representation. The second one is the renumbering of the players; the payoff matrices are transposed and exchanged.

Isomorphism: An isomorphism from $G = (\phi, H)$ to $G' = (\psi, H')$ is a system of mappings $f = (f_i)_N$ where $f_i$ is a one-to-one mapping from player $i$’s pure strategy set $\phi_i$ in $G$ onto $\psi_{\sigma(i)}$, the pure strategy set of player $\sigma(i)$ in $G'$, such that $\sigma$ is a one-to-one mapping of the player set $N$ in $G$ onto the player set $N'$ in $G'$ and $H'$ satisfies conditions of the form

$$(2.8) \quad H'_{\sigma(i)}(f(\psi)) = \alpha_i H_i(\psi) + \beta_i \quad \text{for all } \psi \in \phi$$

for every $i \in N$ with constants $\alpha_i > 0$ and $\beta_i$. (The notation $f(\psi)$ has been explained above.)

Extension to mixed strategies: Consider an isomorphism $f = (f_i)_N$ from $G = (\phi, H)$ to $G' = (\psi, H')$. We write $q'_o(i) = f_i(q_i)$ if we have
(2.9) \( q'_{\sigma(i)}(f_i(q_i)) = q_i(q_i) \)
for every \( q_i \in \Phi_i \). In this way \( f_i \) is extended from \( \Phi_i \) to \( Q_i \). We write \( q' = f(q) \) if the components of \( q' \) are related to those of \( q \) as in (2.9). Obviously (2.8) and (2.9) imply

\[
(2.10) \quad H_{\sigma(i)}(f(q)) = a_i H_i(q) + b_i
\]
for every \( q \in Q \) and every \( i \in N \). It is clear that an isomorphism \( f \) looked upon as a mapping defined on \( Q \) preserves best reply relationships and carries equilibrium points into equilibrium points.

Two games \( G \) and \( G' \) are called isomorphic if at least one isomorphism from \( G \) to \( G' \) exists.

**Invariance with respect to isomorphisms:** A solution function \( L \) for a class of normal form games \( \gamma \) is invariant with respect to isomorphisms if for every isomorphism \( f \) from a game \( G \in \gamma \) to a game \( G' \in \gamma \) (which may or may not be different from \( G \)) we have

\[
(2.11) \quad L(G') = f(L(G))
\]

**Interpretation:** Equation (2.11) is the formal expression of what is meant by saying that isomorphic games should not be treated differently. Invariance with respect to isomorphisms includes invariance with respect to positive linear utility transformations to which it adds an invariance with respect to renaming. A formal description of this latter invariance need not be given here. In our judgement invariance with respect to isomorphisms is an indispensable requirement for any rational theory of equilibrium point selection which is based on strategic considerations exclusively.

With the help of the notion of an isomorphism we can give a precise meaning to the idea that the solution should correctly reflect the symmetries of a game.

**Symmetries:** A symmetry of a game \( G = (\Phi, H) \) is an isomorphism from \( G \) to itself.
Symmetry invariant equilibrium points: An equilibrium point $r$ of $G = (\phi, H)$ is called symmetry invariant if for every symmetry $f$ of $G$ we have $r = f(r)$.

Nash has shown that every finite game in normal form has a symmetry invariant equilibrium point [Nash 1].

A solution function $L$ which is invariant with respect to isomorphisms must assign a symmetry invariant equilibrium point to every game in the class where it is defined:

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\hline
& a & & b & & & \\
\hline
A & 2 & 0 & & 2 & 0 & & \\
\hline
B & 0 & 1 & & 0 & 4 & & \\
\hline
\end{array}
\hspace{1cm}
\begin{array}{c|c|c|c|c|c|c|c|c}
\hline
& a & & b & & & \\
\hline
A & 4 & 1 & & 1 & 0 & & \\
\hline
B & 0 & 2 & & 0 & 2 & & \\
\hline
\end{array}
\hspace{1cm}
\begin{array}{c|c|c|c|c|c|c|c|c}
\hline
& a & & b & & & \\
\hline
A & 2 & 0 & & 2 & 0 & & \\
\hline
B & 0 & 1 & & 0 & 4 & & \\
\hline
\end{array}
\]

renaming of players and strategies
linear transformations of payoffs

$\sigma(1) = 2$ $\sigma(2) = 1$

$f_1(A) = b$ $f_1(B) = a$
f$_2(a) = B$ $f_2(b) = A$

player 1's payoff is divided by 2 and
player 2's payoff is multiplied by 2

Figure II-4: An example of a symmetry

An example of a game with a symmetry is given in figure II-4. The game has three equilibrium points, two in pure strategies, namely $Aa$ and $Bb$ and a mixed one $r = (r_1, r_2)$ with $r_1(A) = 2/3$ and $r_2(s) = 1/3$. The symmetry $f$ carries $Aa$ to $Bb$ and vice versa. The mixed equilibrium point $r$ is the only one which is symmetry invariant. Any solution function $L$ which is invariant with respect to isomorphism cannot assign anything else but $L(G) = r$ to this game.
The payoff vector of \( r \) is \( H(r) = \left( \frac{2}{3}, \frac{1}{3} \right) \). Note that both players receive more at each of both pure strategy equilibrium points. Nevertheless, invariance with respect to isomorphism forces us to adopt \( r \) as the solution. 2/

5. Best reply structure

In the last section we have argued that invariance with respect to positive linear payoff transformations has to be supplemented by invariance with respect to renamings of players and strategies. In this way, we obtained the stronger notion of invariance with respect to isomorphisms.

As we have seen, isomorphisms preserve best reply relationships. One may take the point of view that these relationships contain the essence of a non-cooperative game since no other information is needed in order to determine the set of all equilibrium points. This suggests the idea that two games should be treated in the same way if they do not differ with respect to their best reply relationships. Unfortunately, invariance requirements of this kind turn out to be too strong if they are imposed on the solution function. As we shall see in a later section, one would have to accept counter-intuitive consequences.

Our solution concept is composed of a number of different parts which interact in a process of equilibrium point selection. One of the most important notions which enter the definition of the solution as a building block is that of risk dominance. The concept will be explained in later sections. There we shall argue that an invariance requirement based on best reply considerations is very natural with respect to risk dominance even if it cannot be imposed on the solution concept as a whole.

In order to obtain a clear picture of what constitutes the best reply relationships of a game we shall introduce the definition of a best reply structure. A best reply structure may be thought of as a game form which is even more condensed than the normal form.
For reasons which will become apparent later we must be careful not to eliminate too much in our definition of the best reply structure of a game $G = (\phi, H)$. It is not sufficient to preserve the information on best replies to mixed strategy combinations. It is true that nothing else but a mixed combination can be played in a game but we must also look at all possible expectations a player $i$ may have on the other players' behavior. Player $i$'s expectations may not necessarily take the form of an $i$-incomplete mixed strategy combination.

Player $i$ may think that one of two equilibrium points $U = (U_1, \ldots, U_n)$ and $V = (V_1, \ldots, V_n)$ is known to be the solution by all other players. Either all of them will play their strategies $U_j$ or all of them will play $V_j$. Player $i$ may have subjective probability $z$ for the first alternative and $1-z$ for the second. His expectation can be described as a joint mixture of $U_{-i}$ and $V_{-i}$, symbolically expressed by $zU_{-i} + (1-z)V_{-i}$. If player $i$ holds this expectation he must play a strategy which optimizes against it or, in other words, he must select a best reply to $zU_{-i} + (1-z)V_{-i}$. Therefore, it is necessary to introduce formal definitions of joint mixtures and best replies to them before we can go on to define a best reply structure which covers all possible subjective probability distributions a player may have on the behavior of other players.

All definitions will refer to a fixed game $G = (\phi, H)$ with $\phi = \bigotimes_{i \in \mathcal{N}} \phi_i$.

Joint mixtures: A joint mixture over the $i$-incomplete combinations is a probability distribution $q_{-i}$ over $\phi_{-i}$. The probability assigned to $\phi_{-i}$ by $q_{-i}$ is denoted by $q_{-i}(\phi_{-i})$. We use a dot as a lower index in order to distinguish joint mixtures from combinations of mixed strategies. The set of all joint mixtures over $\phi_{-i}$ is denoted by $Q_{-i}$. 
It is clear that probability distributions over $Q_{-i}$ would yield nothing new. We need not consider more general joint mixtures than those in the sets $Q_i$.

We say that the $i$-incomplete mixed combination $q_{-i}$ generates the joint mixture $q_i$ if for every $\varphi_i = (\varphi_j)_{j \in N \setminus \{i\}}$ we have

$$q_i(\varphi_i) = \prod_{j \in N \setminus \{i\}} q_j(\varphi_j)$$

Every $q_{-i} \in Q_{-i}$ generates a joint mixture $q_i$ but not every $q_i \in Q_i$ is generated by a $q_{-i}$. In this respect 2-person games are special cases since here both $Q_i$ and $Q_{-i}$ coincide with the mixed strategy set $Q_j$ of the other player.

**Hybrid combinations:** A hybrid combination $q_i q_{-i}$ consists of a mixed strategy $q_i \in Q_i$ and a joint mixture $q_{-i} \in Q_{-i}$.

Player $i$'s payoff for $q_i q_{-i}$ is defined as follows:

$$H_i(q_i q_{-i}) = \sum_{\varphi_i \in \Phi_i} \sum_{\varphi_{-i} \in \Phi_{-i}} q_i(\varphi_i) q_{-i}(\varphi_{-i}) H_i(\varphi_i \varphi_{-i})$$

This is player $i$'s subjectively expected payoff if he uses $q_i$ and $q_{-i}$ is his subjective probability distribution on the behavior of the other players. It is clear that $H_i(q_i q_{-i})$ agrees with $H_i(q_i q_{-i})$ if $q_i$ is generated by $q_{-i}$. Instead of (2.13) we can also write

$$H_i(q_i q_{-i}) = \sum_{\varphi_{-i} \in \Phi_{-i}} q_{-i}(\varphi_{-i}) H_i(q_i \varphi_{-i})$$

Payoffs of the other players for $q_i q_{-i}$ could be defined in an analogous way but these payoffs have no theoretical significance.

**Best replies:** $r_i$ is a best reply to $q_{-i}$ if we have

$$H_i(r_i q_{-i}) = \max_{q_i \in Q_i} H_i(q_i q_{-i})$$

The well-known fact that $r_i$ is a best reply if and only if every pure strategy $\varphi_i$ with $r_i(\varphi_i) > 0$ is a best reply holds.
here, too. If we know which are the pure best replies we
have a full overview over all best replies.

r_1 is called a strong best reply to q.i if r_1 is the only
best reply to q.i. A strong best reply must be a pure stra-

gy.

It is clear that r_1 is a best reply to q_1 if and only if r_1
is a best reply to the joint mixture q_1 generated by q_1.

Best reply structure: The set of all pure best replies of
player i to q_1 is denoted by A_i(q_1). The correspondence A_i
which assigns the set A_i(q_1) to q_iεQ_1 is called player i's
best reply correspondence. A = (A_i)_N is the system of best
reply correspondences.

The best reply structure B = (φ,A) of G = (φ,H) consists of
the set of pure strategy combinations φ = X φ_i and the system
A = (A_i)_N of best reply correspondences. i∈N

It is clear that an isomorphism f from G to G' carries the
best reply structure of G to that of G'.

Stability sets: The set of all q_iεQ_1 such that a given pure
strategy φ_i is a best reply to q_i is denoted by S(φ_i). The
set S(φ_i) is called the stability set of φ_i. Obviously S(φ_i)
is the set of all q_i with φ_iεA_i(q_i). One may look upon S as
a correspondence from the union of all φ_i to the union of all
Q_i. In a sense the correspondence S is the inverse of the
system A of best reply correspondences. The pair (φ,S) could
also serve as a formal description of the best reply structure.

Graphical representation for 2x2-games: The best reply structure
of 2x2-games can be visualized with the help of a simple gra-
phical representation. Consider the class of 2x2-games descri-
ed by figure II-5. These games have strong equilibrium points
in the upper left and lower right corners. It is convenient
to introduce the notation u_i and v_i for the losses faced by
player i if he deviates from the equilibrium point U = u_1u_2
and V = v_1v_2, respectively, whereas the other player plays
his equilibrium strategy (see figure II-5).

<table>
<thead>
<tr>
<th>U₂</th>
<th>V₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁₁</td>
<td>a₁₂</td>
</tr>
<tr>
<td>b₁₁</td>
<td>b₁₂</td>
</tr>
<tr>
<td>a₂₁</td>
<td>a₂₂</td>
</tr>
<tr>
<td>b₂₁</td>
<td>b₂₂</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
  u_1 &= a_{11} - a_{21} > 0 \\
  u_2 &= b_{11} - b_{12} > 0 \\
  v_1 &= a_{22} - a_{12} > 0 \\
  v_2 &= b_{22} - b_{21} > 0
\end{align*}
\]

**Figure II-5**: 2x2-games with strong equilibrium point in north-west and south-east corners

A mixed strategy \( q_i \) in a 2x2-game is fully described by one of both probabilities. We shall use the notation

\[
(2.16) \quad p_i = q_i(V_i)
\]

Player 1's strategy \( U_1 \) is a best reply for

\[
(2.17) \quad a_{11}p_2 + a_{12}(1-p_2) \geq a_{21}p_2 + a_{22}(1-p_2)
\]

and \( V_1 \) is his best reply for

\[
(2.18) \quad a_{11}p_2 + a_{12}(1-p_2) \leq a_{21}p_2 + a_{22}(1-p_2)
\]

This yields

\[
(2.19) \quad U_1 \in A_1(q_2) \quad \text{for} \quad 0 \leq p_2 \leq \frac{u_1}{u_1 + v_1}
\]

\[
(2.20) \quad V_1 \in A_1(q_2) \quad \text{for} \quad \frac{u_1}{u_1 + v_1} \leq p_2 \leq 1
\]
Similarly we obtain

\[(2.21) \quad U_2 \epsilon A_2(q_1) \quad \text{for} \quad 0 \leq p_1 \leq \frac{u_2}{u_2+v_2}\]

\[(2.22) \quad V_2 \epsilon A_2(q_1) \quad \text{for} \quad \frac{u_2}{u_2+v_2} \leq p_1 \leq 1\]

We can draw a diagram which represents all mixed strategy combinations as points \((p_1, p_2)\) in a rectangular coordinate system. This is done in figure II-6 for a special case \((u_1 = 2, u_2 = 6, v_1 = 8, v_2 = 4)\). The diagram will be called the stability diagram of the game.

![Diagram](image)

**Figure II-6:** Stability diagram of the game of figure II-5
The regions where the four pure strategy combinations are best replies are indicated in figure II-6. We call these regions the stability regions of the respective pure strategy combinations.

The stability regions are closed rectangles, all of which have one corner in common, the mixed equilibrium point with \( p_1 = u_2/(u_2 + v_2) \) and \( p_2 = u_1/(u_1 + v_1) \). The equilibrium points \( U \) and \( V \) belong to their stability region but the "cross combinations" \( U_1V_2 \) and \( V_1U_2 \) belong to the stability region of the opposite cross combination.

It is interesting to note that the best reply structure of a game in the class of figure II-5 does not depend on anything else but the ratios \( u_1/v_1 \) and \( u_2/v_2 \) of the players' deviation losses at both strong equilibrium points. Absolute payoff levels do not matter. Only ratios of payoff differences are important.

Payoff transformations which preserve the best reply structure: Let \( G = (\phi, H) \) be a game and let \( \psi_j \) be a fixed \( j \)-incomplete pure strategy combination for \( G \). We construct a new game \( G' = (\phi, H') \) with the same set \( \phi \) of pure strategy combinations. For \( i \neq j \) define

\[
(2.23) \quad H'_i(\phi) = H_i(\phi) \quad \text{for every } \phi \in \phi
\]

Let \( \lambda \) be a constant. Player \( j \)'s payoff is defined as follows:

\[
(2.24) \quad H'_j(\phi_j(\psi_j \psi_{-j})) = H_j(\psi_j \psi_{-j}) + \lambda
\]

\[
(2.25) \quad H'_j(\phi_j \psi_{-j}) = H_j(\phi_j \psi_{-j}) \quad \text{for } \psi_{-j} \neq \psi_{-j}
\]

We say that \( G' \) results from \( G \) by adding \( \lambda \) to player \( j \)'s payoff at \( \psi_{-j} \).

A look at (2.14) shows that the same amount \( \lambda q_j(\psi_{-j}) \) is added to every payoff of the form \( H_j(q_j q_{-j}) \) in the transition from \( H_j \) to \( H'_j \). Therefore (2.15) holds in \( G' \) if and only if it holds in \( G \). We obtain the following
result: Adding \( \lambda \) to player \( j \)'s payoff at \( v_{-j} \) does not change the best reply structure.

Consider the games of figure II-5. We receive the game of figure II-7 if we make the following changes one after the other:

1. We add \(-a_{21}\) to player 1's payoffs at \( U_2 \)
2. We add \(-b_{12}\) to player 2's payoffs at \( U_1 \)
3. We add \(-a_{12}\) to player 1's payoffs at \( V_2 \)
4. We add \(-b_{21}\) to player 2's payoff at \( V_1 \)

This confirms once more what we already know from the investigation of the best reply structure of the games of figure II-5: Every game in this class has the same best reply structure as the corresponding game of figure II-7.

It may be worth-while to point out that not every payoff transformation which preserves the best reply structure can be obtained by a combination of positive linear payoff transformations with the repeated application of the operation of adding a constant to player \( j \)'s payoffs at \( v_{-j} \). 2x2-games are exceptional in this respect. Already in 2x3-games other best reply structure preserving payoff transformations are possible.

\[\begin{array}{cc|c|c}
& U_2 & V_2 \\
\hline
U_1 & u_1 & 0 & u_2 & 0 \\
V_1 & 0 & v_1 & 0 & v_2 \\
\end{array}\]

**Figure II-7:** Games received by best reply structure preserving transformations from those of figure II-5
An example is the class of games in figure II-8. A positive linear transformation or adding a constant at player 2's payoffs at a or b cannot change the quotient

\[
\frac{H_2(bd) - H_2(bc)}{H_2(be) - H_2(bd)} = \frac{1+t}{1-t}
\]

which clearly depends on t. Therefore, a combination of such transformations cannot yield the same result as a transition from one t to another.

<table>
<thead>
<tr>
<th></th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1-t</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1+t</td>
<td>(\frac{4}{3} + \frac{2}{3}t)</td>
</tr>
</tbody>
</table>

\(0 < t < 1\)

c is best reply for \(0 \leq q_1(b) \leq \frac{1}{2}\)
d is best reply for \(\frac{1}{2} \leq q_1(b) \leq \frac{3}{4}\)
e is best reply for \(\frac{3}{4} \leq q_1(b) \leq 1\)

Figure II-8: A class of 2x3-games with the same best reply structure

Invariance with respect to payoff transformations which preserve the best reply structure: A solution function \(L\) for a class of games \(G\) is called invariant with respect to payoff transformations which preserve the best reply structure or shortly best reply invariant if for any two games \(G = (\Phi, H)\) and \(G' = (\Phi', H')\) in \(G\) with the same best reply structure we have \(L(G) = L(G')\).
Comment: As has been said before we do not insist on best reply invariance as a desirable property of a solution function. Nevertheless, it is an intuitively attractive requirement which should not be violated without a good reason. We want to keep as much of it as possible.

6. Payoff dominance

Consider the game of figure II-9. The equilibrium point \( U = U_1U_2 \) yields higher payoffs for both players than the other pure strategy equilibrium point \( V = V_1V_2 \). The mixed equilibrium point with probabilities of .4 and .8 for \( U_1 \) and \( U_2 \), respectively, yields even worse payoffs, namely 7.2 for player 1 and 4 for player 2. Clearly, among the three equilibrium points of the game \( V_1V_2 \) is the most attractive one for both players. This suggests that they should not have any trouble to coordinate their expectations at the commonly preferred equilibrium point \( V_1V_2 \). The solution of the game should be \( V_1V_2 \). The idea that equilibrium points with greater payoffs for all players should be given preference in problems of equilibrium point selection leads to the following definition.

**Payoff dominance:** Let \( r \) and \( s \) be two equilibrium points of \( G = (\phi,H) \) with \( \phi = X \phi \). We say that \( r \) **payoff dominates** \( s \) if we have

\[
(2.27) \quad H_i(r) > H_i(s) \quad \text{for every } i \in \mathbb{N}
\]

In (2.27) we require strict inequality since we want to restrict considerations of payoff dominance to cases where the interest of all players unambiguously points in the same direction.

The idea of payoff dominance must be handled with care. We cannot require that \( L(G) \) should never be payoff dominated by any other equilibrium point. As we have seen in section 4 invariances with respect to isomorphisms forces us to accept the mixed equilibrium point as the solutions of the game in figure II-4 even if it is payoff dominated by both pure strategy equilibrium points.
Figure II-9: Example of a 2x2-game with payoff dominance

$$
\begin{array}{c|cc}
U_2 & V_2 \\
\hline
9 & 0 \\
7 & 1 \\
7 & 8 \\
2 & 6 \\
\end{array}
$$

Figure II-10: Game with the best reply structure of the game in figure II-9

$$
\begin{array}{c|cc}
U_2 & V_2 \\
\hline
2 & 0 \\
6 & 0 \\
0 & 8 \\
0 & 4 \\
\end{array}
$$

The example of figure II-4 shows that we should not pay any attention to payoff dominance relationships where the dominating equilibrium point fails to be symmetry invariant. This leads to the following definitions.

**Payoff efficiency:** A symmetry invariant equilibrium point $r$ of a game $G = (\phi, H)$ is called **payoff efficient** if $G$ has no other symmetry invariant equilibrium point $s$ which payoff dominates $r$.

A solution function $L$ for a class of games $\mathcal{G}$ is **payoff efficient** if $L(G)$ is payoff efficient for every $G \in \mathcal{G}$. 
Unfortunately, payoff efficiency is a very strong requirement which cannot be easily satisfied by a solution concept such as ours. Moreover, there are reasons why it should not be satisfied in general. One of these reasons will be discussed in the section on cells.

Another reason is connected to the fact that a situation similar to that in figure II-4 may arise without any lack of symmetry invariance. Two equilibrium points which both payoff dominate a third one but not each other may be equally strong in the sense that the theory does not yield a sufficient reason to select one rather than the other. In such situations it may be unavoidable to select an equilibrium point which fails to be payoff efficient.

In spite of the difficulties arising with this notion, payoff dominance is an important criterion of equilibrium point selection which cannot be completely ignored.

Payoff dominance relationships can easily be reversed by repeated additions of constants to a player j's payoff at some $\psi_j$. Any strong equilibrium point $\psi$ can be made the only payoff efficient one by performing the operations of adding a sufficiently great constant $\lambda_j$ to the payoffs of every player j at his $j$-incomplete $\psi_j$ derived from $\psi$. This shows that best reply invariance and payoff efficiency are in conflict.

In the construction of our solution concept we have rejected full best reply invariance in favor of keeping the possibility of giving some room to considerations of payoff dominance without going as far as imposing the requirement of payoff efficiency.

7. The intuitive notion of risk dominance

Consider the game of figure II-11. There is no payoff dominance relationship between both pure strategy equilibrium points $U = (U_1, U_2)$ and $V = (V_1, V_2)$. Player 1 has higher payoffs at $U$ and player 2 has higher payoffs at $V$. 
Suppose that the players are in a state of mind where they think that either U or V must be the solution of the game. What is the risk of deciding one way or the other? If player 1 expects that player 2 will choose U₂ with a probability of more than .01 it is better for him to choose U₁. Only if player 2 chooses V₂ with a probability of at least .99 player 1's strategy V₁ will be the more profitable one. In this sense U₁ is much less risky than V₁.

```
<table>
<thead>
<tr>
<th></th>
<th>U₂</th>
<th>V₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>U₁</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>V₁</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>51</td>
</tr>
</tbody>
</table>
```

Figure II-11: An extreme example of risk dominance

Now let us look at the situation of player 2. His strategy V₂ is the better one if he expects player 1 to select V₁ with a probability of more than .49 and U₂ is preferable if he expects U₁ with a probability greater than .51. In terms of those numbers V₂ seems to be slightly less risky than U₂.

It is obvious that player 1's reason to select U₁ rather than V₁ is much stronger than player 2's reason to select V₂ rather than U₂. The players must take this into account when they try to form subjective probabilities on the other player's behavior. Presumably player 1 will select U₁ with high probability and since player 2 knows this he is likely to think that it is better for him to choose U₂ rather than V₂. It is plausible to assume that at the end both players will come to the conclusion that both of them will play the equilibrium point U.
The same line of reasoning can be followed for less extreme situations. Consider a game of the form of figure II-7 with $u_1 > v_1$ and $v_2 > u_2$. Player 1's risk situation is connected to the ratio $u_1/v_1$ and player 2's risk situation to the ratio $v_2/u_2$. Player 1 is more strongly attracted to U than player 2 to V if $u_1/v_1$ is greater than $v_2/u_2$. This is the case if and only if we have $u_1 u_2 > v_1 v_2$.

These considerations suggest the following notion of risk dominance for the games under consideration. U risk dominates V for $u_1 u_2 > v_1 v_2$ and V risk dominates U for $v_1 v_2 > u_1 u_2$.

The heuristic arguments which lead to this conclusion are fully in terms of the best reply structure. We have compared probabilities of the form $u_i/(u_i + v_i)$ and $v_i/(u_i + v_i)$. The probabilities which must be compared are the same in the more general situation of figure II-5. These probabilities depend only on the best reply structure.

Since similar products appear in Nash's cooperative bargaining theory we call $u_1 u_2$ and $v_1 v_2$ the Nash-products of U and V, respectively.

It is interesting to note that the areas of the stability regions of U and V (see figure II-5) are proportional to the Nash-products of U and V. This is a further argument for a notion of risk dominance based on the comparison of Nash-products.

Risk dominance and payoff dominance may point in different directions. An example is the game of figure II-9 where U payoff dominates V but V has the greater Nash-product (the Nash-products are the same as in figure II-10).

The notion of risk dominance between strong equilibrium points which has been obtained heuristically can be characterized by a set of simple axioms. This will be done in a later section.
8. Payoff Monotonicity

Consider a game \( G = (\phi, H) \) with \( \phi = \bigcap_{i \in N} \phi_i \) and let \( \phi \) be a pure strategy equilibrium point of \( G \). We construct a new game \( G' = (\psi, H') \) with the same set \( \phi \) of pure strategy combinations. Let \( \lambda_i \) with \( i \in N \) be non-negative constants at least one of which is positive. Define

\[
(2.28) \quad H'(\phi) = H(\phi) \quad \text{for} \quad \phi \neq \psi
\]

\[
(2.29) \quad H'_i(\phi) = H_i(\phi) + \lambda_i \quad \text{for every} \quad i \in N
\]

If \( G \) and \( G' \) are related in this way we say that \( G' \) results from \( G \) by strengthening \( \psi \). The only difference between \( G \) and \( G' \) consists in the fact that some players receive more at \( \psi \).

**Payoff monotonicity**: A solution function \( L \) for a class of normal form games is called payoff monotoneous if the following is true: If the solution \( L(G) \) of a game \( G \in \mathcal{F} \) is a pure strategy equilibrium point and if \( G' \) results from \( G \) by strengthening \( L(G) \) then we have \( L(G') = L(G) \).

**Interpretation**: The requirement of payoff monotonicity is a very appealing one. Why should an equilibrium point become less attractive if some of its payoffs are increased? Nevertheless, an objection can be raised which makes it doubtful whether one should insist on payoff monotonicity as a general property.

In order to explain the nature of the counter-argument we look at the example of the three-person games of figure II-12 and of figure II-13. The game of figure II-13 results from that of figure II-12 by strengthening \( U = U_1U_2U_3 \). In the second game player 3 receives 1 unit more at \( U \) than in the first one. Otherwise both games agree in all payoffs.

It is reasonable to start a crude analysis of the risk situation in both games with the assumption that player 3 is more
Figure II-12: A three-person game. Player 3 chooses between the left and the right matrix.

Figure II-13: A game which results from that of figure II-12 by strengthening $U$.

likely to choose $U_3$ in the second game. But does this strengthen $U$ more than $V = V_1V_2V_3$?

Suppose that each of the players 1 and 2 expects the other to behave in the same way in both games. Then an increase of their subjective probability for $U_3$ will increase their incentive to use their strategies $V_1$ and $V_2$. The numbers are chosen in such a way that it is not unreasonable to expect that the change from the first game to the second one enhances the stability of $V$ more than that of $U$. 
The solution concept which we shall propose here actually assigns the solution \( U \) to the first game and the solution \( V \) to the second. It does not have the payoff monotonicity property.

In spite of the fact that we reject payoff monotonicity as a general property we think that it is a very reasonable requirement for 2x2-games. There we cannot find any reason to suppose that one of two strong equilibrium points can be made more attractive than the other. The nature of the example seems to indicate that at least three players are needed in order to produce an example where payoff monotonicity fails to be convincing.

9. Axiomatic characterization of risk dominance between strong equilibrium points in 2x2-games

Let \( \mathcal{G} \) be the class of all 2x2-games with 2 strong equilibrium points. We shall axiomatise a risk dominance relationship which is defined between the two strong equilibrium points of any game in \( \mathcal{G} \). The notation \( U \vdash V \) is used in order to indicate that \( U \) risk dominates \( V \). We also permit that neither \( U \) risk dominates \( V \) nor \( V \) risk dominates \( U \) and we write \( U \| V \) if this is the case. For any game \( G \in \mathcal{G} \) with strong equilibrium points \( U \) and \( V \) exactly one of the following statements must hold:

1. \( U \vdash V \) \( U \) risk dominates \( V \) in \( G \)
2. \( V \vdash U \) \( V \) risk dominates \( U \) in \( G \)
3. \( U \| V \) There is no risk dominance between \( U \) and \( V \) in \( G \).

This is part of the definition of the concept of a risk dominance relationship and not yet a requirement to be imposed on it.

The axioms are stated below. It will always be understood that \( U \) and \( V \) are the strong equilibrium points of a game \( G = (\phi,H) \in \mathcal{G} \).
(I). Invariance with respect to isomorphisms: Let $f$ be an isomorphism from $G$ to $G'$. Then we have $f(U) \preceq f(V)$ in $G'$ if and only if we have $U \preceq V$ in $G$.

(II). Best reply invariance: Let $G' = (\phi', H')$ be a game which has the same best reply structure as $G = (\phi, H)$. Then $U \preceq V$ holds in $G'$ if and only if it holds in $G$.

(III). Payoff monotonicity: Let $G' = (\phi', H')$ be a game which results from $G = (\phi, H)$ by strengthening $U$. If $U \preceq V$ or $U \mid V$ holds in $G$ then $U \preceq V$ holds in $G'$.

Interpretation: It is clear that we must require invariance with respect to isomorphisms. The reasons are the same as those discussed in section 4. As we have seen in section 7, the intuitive arguments which we have used in order to compare risks attached to different equilibrium points run in terms of the best reply structure. Imposing axiom (II) means that we look for a concept of this kind without specifying a precise way in which risk comparisons should be made.

Payoff monotonicity has been discussed in section 8. As far as $2 \times 2$-games are concerned it seems to be a very desirable property even if for more complicated games the situation is less clear.

Theorem: There is one and only one risk dominance relationship for $\mathcal{G}$ which satisfies (I), (II) and (III). As in figure II-5 let $u_i$ and $v_i$ with $i = 1, 2$ be the deviation losses of player $i$ at the strong equilibrium points $U$ and $V$ of a game $G \in \mathcal{G}$. Then we have

\begin{align*}
(2.30) \quad U & \preceq V \quad \text{for } u_1u_2 > v_1v_2 \\
(2.31) \quad V & \preceq U \quad \text{for } v_1v_2 > u_1u_2 \\
(2.32) \quad U \mid V \quad \text{for } u_1u_2 = v_1v_2
\end{align*}
Proof: Up to renamings of the strategies every game $G \in \mathcal{X}$ is in the class of games of figure II-5. Any such game has the same best reply structure as the corresponding game of figure II-7 (see section 5). Multiplication of player 1's payoff by $1/v_1$ and player 2's payoff by $1/u_2$ transforms a game of figure II-7 into a game of figure II-14.

\[
\begin{array}{c|cc}
  & V_1 & V_2 \\
\hline
U_1 & u & 0 \\
   & 1 & 0 \\
U_2 & 0 & 1
\end{array}
\]

\[u = \frac{u_1}{v_1}, \quad v = \frac{v_2}{u_2}\]

Figure II-14: Games equivalent to those of figure II-7

For $u = v$ the game of figure II-14 has a symmetry which carries $U$ to $V$ (renaming of strategies and exchanging the players). Therefore, in view of (I) for $u = v$ we must have $U \sim V$.

A game of figure II-14 with $u > v$ results from a game with $u = v$ from strengthening $U$. Therefore, in view of (III) we must have $U \succ V$ for every game of figure II-14 with $u > v$ and similarly $V \succ U$ for every game of figure II-14 with $v > u$.

Since the best reply structure of a game of figure II-5 is the same as that of the corresponding game of figure II-14 we must have $U \succ V$ for $u > v$ there, too. We have $u > v$ if and only if $u_1v_2 > v_1u_2$. Analogously, we have $V \succ U$ if and only if $v_1v_2 > u_1u_2$. This proves the theorem.

Comment: The theorem gives a firm basis to our intuitive considerations on risk dominance between strong equilibrium points in $2 \times 2$-games. The only notion of risk dominance which
agrees with the axioms can be described as a comparison of Nash-products of deviation losses.

It is interesting that our result supports Nash's bargaining theory under fixed threats without relying on anything similar to the axiom of irrelevant alternatives which plays a crucial role in his axiomization.

On the basis of the risk dominance relationship characterized by the theorem one can define a solution function which will be called pure risk dominance solution function since it completely ignores the aspect of payoff dominance.

The pure risk dominance solution: The pure risk dominance solution function \( \tilde{L} \) on \( \mathcal{A} \) is defined as follows: Let \( U \) and \( V \) be the strong equilibrium points of \( G = (\varphi, H) \) and let \( u_i \) and \( v_i \) for \( i = 1, 2 \) be the deviation losses at \( U \) and \( V \) (as in Figure II-5). Let \( r = (r_1, r_2) \) with

\[
(2.33) \quad r_1(U_1) = \frac{v_2}{u_2 + v_2}, \quad r_2(U_2) = \frac{v_1}{u_1 + v_1}
\]

be the third equilibrium point of \( G \). Then we have:

\[
(2.34) \quad \tilde{L}(G) = \begin{cases} 
U & \text{for } u_1u_2 > v_1v_2 \\
V & \text{for } v_1v_2 > u_1u_2 \\
r & \text{for } u_1u_2 = v_1v_2
\end{cases}
\]

Conflict between risk dominance and payoff dominance: We have already pointed out in section 7 that a risk dominance relationship in one direction is compatible with a payoff dominance relationship in the other direction. It is maybe useful to look at the extreme example of figure II-15.

Here \( U \) payoff dominates \( V \) but \( V \) strongly risk dominates \( U \). It is reasonable to expect that most players would prefer to play \( V_1 \) rather than \( U \) if the game is played for a considerable amount of money (say $1000,- per unit) without preplay communication. On the other hand, with preplay com-
Figure II-15: Example of payoff dominance and risk dominance in opposite directions

munication they may very well come to the conclusion that they
can trust each other to choose \( U = (U_1, U_2) \). An agreement
to do so is selfstabilizing and does not need any commit-
ment power.

If it is common knowledge of both players that both are
fully rational then there should not be any need to enter
preplay communication before the beginning of this game
since the outcome can be predicted easily anyhow. There-
fore, even under conditions which do not permit preplay
communication they should trust each other to play \( U \).

The pure risk dominance solution involves a certain lack
of rationality. Nevertheless, under certain circumstances
distrust may be justified. Suppose for example that in the
game under consideration preplay communication has taken
place and for some mysterious reason the players could not
agree on \( U \). Then, after the breakdown of communication, it
is certainly justified not to look at payoff dominance and
to rely on risk dominance only.

For a long time the authors took the point of view that
everything which goes beyond pure risk dominance should be
captured by formal models of preplay communication which
explicitly describe how trust is developed rationally under
the threat of conflict. In a theory of this type the pure
risk dominance solution would serve as a threat point of preplay bargaining. Preplay bargaining itself would be described as a game where an equilibrium point has to be selected. Hopefully in this bargaining game the conflict between risk dominance and payoff dominance may not occur. Otherwise one would meet the difficulty that bargaining on bargaining is required before the beginning of the bargaining game. In spite of the difficulties involved in this approach it may still be worth trying.

It is our impression that a theory which gives room to both payoff dominance and risk dominance is more in agreement with the usual image of what constitutes rational behavior. Moreover, it avoids some of the difficulties of the approach outlined above even if models of preplay communication may still be necessary for some purposes.

10. The proposed solution function for 2x2-games with two strong equilibrium points

The solution function \( L \) for \( \mathcal{G} \) which results from the application of our general concept to this class gives absolute priority to payoff dominance. It can be described as follows. Let \( U \) and \( V \) be the strong equilibrium points of \( G = (\Phi,H) \). Then we have:

\[
L(G) = \begin{cases} 
U & \text{if } H_i(U) > H_i(V) \text{ for } i = 1,2 \\
V & \text{if } H_i(U) < H_i(V) \text{ for } i = 1,2 \\
I(G) & \text{else}
\end{cases}
\]

(2.35)

where \( I(G) \) is the pure risk dominance solution function introduced in section 9. We call this solution function \( L \) the proposed solution function for \( \mathcal{G} \).

One may ask how the solution function \( L \) should be extended to the class of all 2x2-games. Obviously, those games which have only one equilibrium point raise no difficulties. Some degenerate cases with an infinity of equilibrium points like the example of figure 11-16 cannot be fully discussed before the introduction of further basic concepts. An important de-
Figure II-16: A degenerate 2x2-game

<table>
<thead>
<tr>
<th></th>
<th>U₂</th>
<th>V₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>U₁</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>V₁</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Definition, namely that of a cell will be introduced in the next section in order to prepare the impossibility result on sequential player splitting which has been mentioned in the introduction of the chapter.

II. Cells

It is natural to require that a solution function for extensive games is subgame consistent in the sense that the behavior prescribed on a subgame is nothing else than the solution of the subgame. After all, once the subgame has been reached all other parts of the game are strategically irrelevant.

It is not immediately clear how subgame consistency can be achieved in the framework of the agent normal form. The definition of a subgame depends on the tree structure of the extensive form. The agent normal form abstracts from the information on the sequential order in which choices are made.

Nevertheless, the essential features of a subgame are not lost in the transition to the agent normal form. In order to capture these essential features we shall define substructures of the agent normal form which will be called cells. As we shall see, a subgame always corresponds to a cell but it is also possible that an agent normal form has
a cell which does not arise from a subgame of the extensive form. Concentration on the essential features of a subgame leads us to the more general notion of a cell.

Before we proceed to the definition of a cell we shall investigate the question what happens to a subgame in the transition to the agent normal form. This shall be done with the help of an example. After the discussion of the example it will be easy to see the general picture.

Subgames and the agent normal form: Let $\Gamma$ be the extensive game of figure II-17. The game $\Gamma$ has two proper subgames, one following the choice $\xi$ of player 1 and another after his choice $r$. The former subgame will be called $\Gamma_\xi$ and the latter $\Gamma_r$.

What happens to $\Gamma_\xi$ and $\Gamma_r$ in the agent normal form of $\Gamma$? In the agent normal form each information set $u_i$ belongs to a separate player $i$. Thus the agent normal form has two new players 2 and 4 which control $u_2$ and $u_4$ instead of the old player 2. Similarly the old player 3 is split into two new players 3 and 5.

Obviously, a subgame can be looked upon as a game which is played by a subset of all agents. Thus $\Gamma_r$ is played by the new players 4 and 5. Once the subgame has been reached, the payoffs depend only on the behavior of the agents in the subgame. In this sense the agents of a subgame depend only on each other and not on the other agents of the game. As far the agents 4 and 5 are concerned the strategic situation is that of the game of figure II-18.

The fact that the subgame agents are independent of outside agents is less obvious if one looks at the agent normal form without knowing from which extensive form it is derived. Since an agent always receives the payoff of the corresponding player in the original game, he receives payoffs not only inside the subgame but also at other endpoints. Moreover, outside agents decide with which probability the subgame is reached. In the agent normal form of the game of figure II-17 let $p_i$ be the probability with which $r$ is chosen by player 1. Then each of the agents 4 and 5 re-
Figure II-17: Example of an extensive game with two proper subgames. Information sets are represented by dashed lines. Choices are indicated by the letters l and r (standing for left and right). Payoff vectors are indicated by column vectors above the corresponding endpoints.

eives $p_1$ times his payoff in the subgame $\Gamma_r$ plus $1-p_1$ times his payoff in the subgame $\Gamma_l$. For fixed choices of agents 1, 2, and 3 the payoffs of players 4 and 5 are non-negative linear transformations of the payoffs of figure II-18. The transformation is non-negative but not necessarily positive since the coefficient $p_1$ may be zero.
Figure II-18: The strategic situation of agents 4 and 5 in the game of figure II-17

If \( p_1 \) is zero then \( r_r \) is not reached. In the perturbed agent normal form \( p_1 \) is constrained by a positive lower bound and the transformation is always positive.

It can be seen easily that the emerging picture holds for subgames in general. Let \( C \) be the set of agents in a subgame and let \( N \) be the set of all agents. Then the agents in \( C \) are independent of outside agents in the sense that up to non-negative linear transformations their payoffs depend on their choices only. Moreover, in a perturbed agent normal form these transformations are always strictly positive since the multiplicative coefficient is the probability that the subgame is reached. The additive constant is due to payoffs at outside endpoints.

One may say that in the perturbed game a change of the outside agent's strategies has essentially the same effect on the agents in \( C \) as a transition to an equivalent game (see section 3). This is the distinguishing feature of a subgame which shall be captured by our definition of a cell.

In the case of a subgame, the multiplicative coefficient of the linear transformation connected to a strategy change by outside agents is always the same for all agents of the subgame. There seems to be no good reason why this peculia-
rity should be reproduced by the cell notion. Therefore, in our definition of a cell we shall permit different multiplicative constants for different agents.

Suppose that two players 1 and 2 with linear utilities in money are involved in a bimatrix game whose entries are in terms of unknown currencies. Before they make their choices a third player secretly selects between two alternative possibilities (a) player 1 receives dollars and player 2 receives Israeli pounds or (b) player 1 receives French francs and player 2 receives German marks. - It is reasonable to define cells in such a way that in this example players 1 and 2 form a cell. Obviously, their strategic situation is not influenced by the currency assignment.

We may change the currency example by giving an additional choice (c) to the third player. If he selects (c) then player 1 and 2 will receive zero payoffs no matter what they do. We shall take the point of view that 1 and 2 form a cell in this case, too. In the perturbed games 1 and 2 would form a cell anyhow, even if a definition were adopted which would be based on positive transformations only. It seems to be preferable to work with a cell concept which does not give different results for an agent normal form and its perturbed agent normal forms. Therefore, we shall permit zero as a multiplicative coefficient.

In order to prepare the definition of a cell we must introduce some auxiliary definitions and notations which all refer to a fixed game \(G = (\phi, H)\) with \(\phi = \sum_{i \in N} \phi_i\).

Subsets strategy combinations: Let \(C\) be a subset of \(N\). A subset strategy combination for \(C\) or shortly a strategy combination for \(C\) is a collection \(q_C = (q_i)_C\) which contains a strategy \(q_i \in Q_i\) for every \(i \in C\). A pure strategy combination \(\phi_C = (\phi_i)_C\) for \(C\) contains a pure strategy \(\phi_i \in \phi_i\) for every \(i \in C\). The set of all pure strategy combinations for \(C\) is denoted by \(\phi_C\) and \(Q_C\) stands for the set of all mixed strategy combinations for \(C\). Instead of \(\phi_{N \setminus C}\) and \(q_{N \setminus C}\) we also use the shorter notation
\( \phi_c \) and \( q_c \). In \( \phi_c \) and \( q_c \), too, \( \neg C \) stands for \( N \neg C \).

**Fixed players:** Let \( r_c \) be a subset strategy combination for \( N \neg C \). We construct a game \( G' = (\phi', H') \) with

\[
(2.36) \quad \phi' = \bigwedge_{i \in C} \psi_i
\]

The payoff function \( H' \) is defined as follows:

\[
(2.37) \quad H_i'(\psi_i) = H_i(r_c \psi_C) \quad \text{for } i \in C
\]

We say that \( G' \) is the game which results from \( G \) by fixing the players in \( N \neg C \) at \( r_c \).

**Centroids:** Let \( D \) be a non-empty subset of \( N \) and for every \( i \in D \) let \( \psi_i \) be a non-empty subset of this player's pure strategy set \( \psi_i \). The **centroid** of \( \psi = \bigwedge_{i \in D} \psi_i \) is a strategy combination \( r_D = (r_i)_D \) for \( D \) which is defined as follows:

\[
(2.38) \quad r_i(\psi_i) = \begin{cases} 
1/|\psi_i| & \text{for } \psi_i \in \psi_i \\
0 & \text{for } \psi_i \notin \psi_i 
\end{cases}
\]

for every \( i \in D \) where \(|\psi_i|\) is the number of elements in \( \psi_i \).

The centroid of \( \psi \) is denoted by \( c(\psi) \).

**Cells:** Let \( C \) be a non-empty proper subset of \( N \) and let \( G^C = (\phi^C, H^C) \) be the game which results from \( G \) by fixing the players in \( N \neg C \) at the centroid \( C(\phi_c) \) of the cartesian product of their pure strategy sets. The game \( G^C \) is a **cell** of \( G \) if for every \( \psi_c \in \phi_c \) and for every \( i \in C \) a number \( \alpha_i(\psi_c) > 0 \) and number \( \beta_i(\psi_c) \) can be found such that we have:

\[
(2.39) \quad H_i(\psi_c \psi_C) = \alpha_i(\phi_c) H_i^C(\psi_c) + \beta_i(\psi_c)
\]

We say that \( C \) forms a **cell** if \( G^C \) is a cell. If this is the case \( G^C \) is the cell formed by \( C \).

**Remark:** If \( G \) is the agent normal form of an extensive game and \( C \) is the set of all agents in a subgame then \( C \) forms a cell in \( G \). This is clear from the discussion in the sub-
section on subgames and the agent normal form. On the other
hand, a cell may not necessarily arise from a subgame.

In the case of a cell arising from a subgame the subgame
is always reached with positive probability by the centroid
c(φ_c). Therefore, the definition of G^c with the help of
the centroid always results in a normal form which is equi-
valent to the agent normal form of the subgame.

The centroid c(φ_c) serves to pick a specific game as a re-
presentation of the class of all games G' which result from
G by fixing the players in N\C at a completely mixed strategy
combination q_c. Here completely mixed means that every player
j∈N\C selects each of his pure strategies with positive pro-
bability. All these games are equivalent. Therefore, the use
of the centroid in the definition of a cell is non-arbitrary.

Lemma on cells: Let C and C' with C ∩ C' ≠ ∅ be two proper
subsets of N which both from cells of G = (φ,ξ) with ξ = ξ_i
∀ i∈N. Then D = C∪C' forms a cell of G, too.

Proof: Any change of the strategy combination q_D for N\D
can be achieved by two successive changes, such that first
only the strategies of players in N\C are changed and then
only those of players in C\D. Both changes are connected
with non-negative linear payoff transformations for the
payoffs of players in D, in the first case since C forms a
Cell, in the second since C' forms a cell. Two successive
non-negative linear transformations performed one after another
are equivalent to non-negative linear transformation. In this
way we receive the non-negative linear transformations whose
existence is required by the definition of a cell as applied
to D.

Counterexample: One may think that the union of two subsets
C and C' form a cell if C and C' form cells and if the union is
a proper subset of N. The example of the game exhibited in fi-
gure II-19 shows that this is not necessarily true. There
both (1) and (2) form cells since for fixed strategies of the
other players the difference between the payoffs for V_1 and U_1
is always 1.
Figure II-19: A counterexample. \{1\} and \{2\} form cells but \{1,2\} does not form a cell. Player 1, 2 and 3 choose rows, columns and matrices, respectively.

The situation is the same one for player 2. Nevertheless, \{1,2\} is not a cell since a shift from \(U_3\) to \(V_3\) reverses the sign of the payoff differences between \(U_1V_2\) and \(V_1V_2\). No non-negative payoff transformation can produce a result like this.

**Elementary cells:** Let \(C\) be a non-empty subset of \(N\) which forms a cell \(G^C\). The cell \(G^C\) is called elementary if no proper subset of \(C\) forms a cell in \(G\). It follows by the lemma on cells that subsets which form elementary cells do not intersect.

**Comment:** The fact that elementary cells do not intersect is an important one since it enables us to define a solution function which is based on the idea that a game with cells should be solved by first solving the elementary cells and then solving the game which results by fixing the players of the elementary cells at the strategies prescribed by the solutions of these cells.

12. Cell consistency and truncation consistency

In this section we shall look at two additional desirable properties of solution functions. Roughly speaking, cell consistency requires that the solution of the whole game agrees
with that of its cells as far as the cell players are concerned. Truncation consistency concerns the "truncated" game which results if the players in a cell are fixed at the solution of this cell. The requirement postulates that the solution of this game should agree with the solution of the original one, as far as its players are concerned.

Completeness: A class \( \mathcal{G} \) of normal form games is called complete if a game \( G' \), which results from a game \( G \in \mathcal{G} \) by fixing some but not all of the players at arbitrary strategies, also belongs to \( \mathcal{G} \).

Truncations: Let \( L \) be a solution function for a complete class \( \mathcal{G} \) of games in normal form. Let \( G \in \mathcal{G} \) be a game with a cell \( G^C \). The truncation of \( G \) with respect to \( G^C \) and \( L \) is the game which results from \( G \) by fixing the players of \( G^C \) at their strategies in the solution \( L(G^C) \) of \( G^C \).

Remark: The completeness of \( \mathcal{G} \) is important since it guarantees that both \( G^C \) and the truncation \( G' \) of \( G \) with respect to \( G^C \) and \( L \) will be in \( \mathcal{G} \), if \( G^C \) is a cell of \( G \). Both \( G^C \) and \( G' \) result from \( G \) by fixing some of the players but not all of them.

Cell consistency: A solution function \( L \) for a complete class \( \mathcal{G} \) of normal form games is called cell consistent if for a cell \( G^C \) of a game \( G \in \mathcal{G} \), the solutions \( L(G^C) \) and \( L(G) \) of \( G^C \) and \( G \) always prescribe the same strategies to the players of \( G^C \).

Truncation consistency: A solution function \( L \) for a complete class \( \mathcal{G} \) of normal form games is called truncation consistent if for a truncation \( G' \) of a game \( G \in \mathcal{G} \) with respect to a cell \( G^C \) of \( G \) and \( L \), the solutions \( L(G') \) and \( L(G) \) always prescribe the same strategies for all players of \( G' \).

Interpretation: As far as their strategic situation is concerned the players in a cell do not depend on outside players. This has been discussed in section 11. Obviously, cell consistency is a very natural requirement.

Truncation consistency is a very natural requirement, too, since the outside players know that the cell players do not depend on them. It is rational to expect that the cell
players will play the cell solution. Therefore, the outside players find themselves in the situation of the truncated game.

As we shall see cell consistency and truncation consistency have the consequence that it is sufficient to know the solutions of the games without cells in order to compute the solutions of all games in a complete class.

Decomposibility: A game $G$ is called decomposable if it has at least one cell. Games without cells are called indecomposable. We say that $G$ is fully decomposable if every player belongs to an elementary cell. Decomposable games which are not fully decomposable are called partially decomposable.

Main truncation: Let $L$ be a solution function for a complete class $\mathcal{G}$. For every partially decomposable game $G \in \mathcal{G}$ we construct a game $G'$ which is called the main truncation of $G$. Let $G^1, \ldots, G^k$ be the elementary cells of $G$. The game $G'$ results from by fixing the players in the elementary cells at their strategies in the solutions $L(G^1), \ldots, L(G^k)$ of the elementary cells.

Composition: Let $L$ be a solution function for a complete class of games $\mathcal{G}$ and let $G \in \mathcal{G}$ be a fully decomposable game. Let $r$ be the strategy combination for $G$ which contains for every player $i$ his strategy prescribed by the solution $L(G^j)$ of the elementary cell to which he belongs. This strategy combination $r$ is called the composition of the elementary cell solutions. Now consider a partially decomposable game $G \in \mathcal{G}$. Let $r$ be the strategy combination which (a) for every player $i$ in an elementary cell $G^j$ of $G$ contains his strategy in $L(G^j)$ and (b) for every player in the main truncation $G'$ of $G$ contains his strategy in $L(G')$. This strategy combination $r$ is called the composition of the main truncation and elementary cell solutions.

Extension: Let $\mathcal{G}$ be a complete class of normal form games and let $\mathcal{G}_0$ be the subclass of all indecomposable games in $\mathcal{G}$. Moreover, let $L_0$ be a solution function for $\mathcal{G}_0$. On the basis of $L_0$ we shall construct a solution function $L$ for $\mathcal{G}$.
which will be called the extension of $L_0$ to $\mathcal{G}$. The extension $L$ is recursively defined by the following properties (A), (B) and (C).

(A) For $G \in \mathcal{G}_0$ we have $L(G) = L_0(G)$

(B) If $G \in \mathcal{G}$ is fully decomposable then $L(G)$ is the composition of the elementary cell solutions.

(C) If $G \in \mathcal{G}$ is partially decomposable then $L(G)$ is the composition of the main truncation and the elementary cell solutions.

It is clear that in this way a solution $L(G)$ is uniquely defined for every game $G \in \mathcal{G}$. Property (C) may have to be applied several times first to the game itself, then to its main truncation, etc. but finally a truncation will arise which is either indecomposable or fully decomposable.

Extension theorem: Let $\mathcal{G}$ be a complete class of games in normal form, let $\mathcal{G}_0$ be the subclass of indecomposable games in $\mathcal{G}$ and let $L_0$ be a solution function for $\mathcal{G}_0$. There is one and only one cell consistent and truncation consistent solution function for $\mathcal{G}$ which agrees with $L_0$ on $\mathcal{G}_0$, namely the extension $L$ of $L_0$ to $\mathcal{G}$.

Proof: It is clear that a cell consistent and truncation consistent solution function must agree with the extension $L$ of $L_0$ since these two properties permit us to compute the solution with the help of (A), (B) and (C).

It remains to show that the extension $L$ of $L_0$ has the properties of cell consistency and truncation consistency. This will be done by induction on the number of players in $C$. Both properties trivially hold for 1-person games. Assume that they hold for games in $\mathcal{G}$ with at most $n-1$ players.

Consider a decomposable game $G \in \mathcal{G}$ with $n$ players; let $G^C$ be a cell of $G$ and let $G'$ be the truncation of $G$ with respect to $G^C$ and $L$. We have to show that $L(G)$ prescribes the same strategies as $L(G')$ and $L(G''')$.
Let $D$ be the set of all players in $G^c$ which do not belong to elementary cells. Let $E$ be the set of all players who belong to elementary cells outside $G^c$. Let $G''$ be the main truncation of $G$.

If $D$ and $E$ are both empty, then $G'$ agrees with $G''$ and the assertion is an immediate consequence of the definition of $L$.

Suppose that we have $D \neq \emptyset$. Let $G^D$ be the main truncation of $G^c$. It is clear that $G^D$ is a cell of $G''$ since the non-negative linear transformations which establish the cell property of $G^D$ in $G'$ can easily be constructed from those which establish the cell property of $G^c$ in $G$.

Suppose that $E \neq \emptyset$. Let $G^1, \ldots, G^m$ be the elementary cells with players in $E$. For $j = 1, \ldots, m$ let $\hat{G}^j$ be the game which results from $G'$ by fixing the players in the cells $G^1, \ldots, G^j$ at their strategies in $L(G^1), \ldots, L(G^j)$. Let $N_1, \ldots, N_m$ be the player sets of $G^1, \ldots, G^m$, respectively. It can be seen easily that for $j = 2, \ldots, m$ the set $N_j$ forms a cell in $G^{j-1}$ even if this cell may not be an elementary one. The non-negative linear transformations which establish the cell property of $N_j$ in $G^{j-1}$ can easily be constructed from those which establish the cell property of $N_j$ in $G$.

Since $G'$ has fewer than $n$ players we can repeatedly apply cell consistency and truncation consistency to the games $G', \hat{G}^1, \ldots, \hat{G}^m$ in order to conclude that $L(G')$ prescribes the strategies in $L(G^1), \ldots, L(G^m)$ and $L(\hat{G}^m)$.

Consider the case $D = \emptyset$ and $E \neq \emptyset$. In this case we have $\hat{G}^m = G''$. This together with our conclusion on $G'$ immediately yields the assertion.

Assume $D \neq \emptyset$ and $E = \emptyset$. In this case $G$ is the truncation of $G''$ with respect to $G^D$ and the assertion follows by the application of cell consistency and subgame consistency to $G''$, $G^D$ and $G'$.

Finally consider the case $D \neq \emptyset$ and $E \neq \emptyset$. Here the assertion follows by the fact that on the one hand $\hat{G}^m$ is related
to \( G' \) in the way which has been explained above and that, on the other hand, \( \tilde{G}^m \) is the truncation of \( G'' \) with respect to \( G^D \). The arguments for the cases \( D = \emptyset, E \neq \emptyset \) and \( D \neq \emptyset, E = \emptyset \) can be combined in order to obtain the result.

Comment: The extension theorem shows that cell consistency and truncation consistency are powerful properties which reduce the task of defining a solution concept to the task of defining a solution concept for indecomposable games.

Cell consistency and truncation consistency require that all considerations which may influence the selection of equilibrium points are applied strictly locally, i.e. only to those indecomposable games which appear in the process of computing the solution with the help of (A), (B) and (C) on the basis of a solution concept for indecomposable games. These indecomposable games shall be called the bricks of the original game.

Local and global payoff efficiency: Payoff efficiency is an example of a selection criterion which cannot be applied to the game as a whole but only locally to its bricks. Figure II-17 shows an example of a conflict between global and local payoff dominance. The subgame \( \Gamma_\ell \) after player 1's choice \( \ell \) has two strong equilibrium points, namely \((\ell, \ell)\) and \((r, r)\). The same is true for the subgame \( \Gamma_r \) after player 1's choice \( r \). The agent normal form of the game lies in no other cells than those corresponding to the subgames.

\((\ell, \ell)\) is the only payoff efficient equilibrium point of the cell corresponding to \( \Gamma_\ell \) and also the only payoff efficient equilibrium point in the cell corresponding to \( \Gamma_r \). (Player 1's agent is not a player in these cells.) If payoff efficiency is applied locally as a selection criterion we must select \((\ell, \ell)\) in both cells. Obviously, if this is done player 1 faces a choice between payoff 4 for \( \ell \) and payoff 3 for \( r \). He has to choose \( \ell \) in the truncated game. Local application of the payoff efficiency criterion yields the equilibrium point where all five agents choose \( \ell \). The payoffs are 4 for everybody.
Consider the strategy combination where all agents choose $r$. This is also an equilibrium point. It yields a payoff of 5 for everybody. It payoff dominates the equilibrium point, where all agents choose $\ell$. It is in the interest of everybody to play this equilibrium point rather than the other one. Unfortunately, this is true only at the beginning of the game. After the subgame $\Gamma_r$ has been reached the interests of player 1 do not count anymore and it is now in the interest of all others to play $(\ell, \ell)$.

Both equilibrium points, that one where all agents choose $\ell$ and that one where all agents choose $r$, are uniformly perfect. It can be seen easily that sufficiently small perturbances do not matter.

13. Sequential agent splitting

Figure II-20 shows what sequential agent splitting means in the extensive form. An agent of player $j$ who has to choose between $a$, $b$ and $c$ is split into two agents such that first one has to select either $a$ or $bc$ and then in case of $bc$ the other decides between $b$ and $c$. In the graphical representation of the extensive form the upper substructure shown by figure II-20 is taken out and the lower one is put in.

At least, at first glance it is hard to imagine why sequential agent splitting should in any way change the strategic situation. Nevertheless, as we shall see, one cannot avoid the conclusion that sequential agent splitting does have a considerable influence on risk comparisons between equilibrium points in some games.

For the purposes of our theory sequential agent splitting must be formally defined in the framework of the agent normal form. Since there the agents are players we shall use the term player splitting instead of agent splitting. Moreover, we shall drop the word sequential since the order in which the decisions are made is not really important in the agent normal form.
Figure II-20: An example of sequential agent splitting in the extensive form

Player splitting: Let $G = (\phi, H)$ with $\phi = \bigotimes_{i \in N} \phi_i$ be a game in normal form; let $j \in N$ be one of the players and let $\phi_j \in \phi_j$ be one of his pure strategies. Moreover, let $k$ be a positive integer with $k \in N$ and define $N' = N \cup \{k\}$. We construct a game $G' = (\phi', H')$ with

\begin{align*}
\phi' &= \bigotimes_{i \in N'} \phi_i^j \\
\phi'_i &= \phi_i \quad \text{for } i \in N \setminus \{j\}
\end{align*}
\[(2.42) \quad \phi' = \{\psi_j, -\psi_j\}_k\]
\[(2.43) \quad \phi_j = \psi_j - \{\psi_j\}\]

(In (2.42) the alternative of not choosing \(\psi_j\) is symbolized by \(-\psi_j\).) We say that \(q' = (q'_i)_N\) corresponds to \(q = (q_i)_N\) and write \(q' \rightarrow q\) if we have

\[(2.44) \quad q_i = q'_i \quad \text{for} \quad i \in \mathbb{N} \setminus \{j\}\]
\[(2.45) \quad q_j(\psi_j) = q'_k(\psi_j)\]
\[(2.46) \quad q_j(\psi_j) = q'_j(-\psi_j)q'_j(\psi_j) \quad \text{for} \quad \psi_j = \psi_j\]

The payoffs for \(G'\) are defined as follows:

\[(2.47) \quad H'_i(\phi') = H_i(\phi) \quad \text{with} \quad \phi' \rightarrow \phi \quad \text{for} \quad i \in \mathbb{N}\]
\[(2.48) \quad H'_k(\phi') = H'_j(\phi') \quad \text{for every} \quad \phi' \in \phi\]

The game \(G' = (\phi', H')\) is called the game which results from \(G = (\phi, H)\) by splitting off a player \(k\) for \(\psi_j\).

**Interpretation:** Even if the formal definition may appear to be somewhat complicated it can be seen easily that it is the correct translation of the idea of sequential agent splitting into the language of the agent normal form. In figure II-20 strategy \(\psi_j\) corresponds to the choice \(a\) and player \(k\) corresponds to the agent who chooses between \(a\) and \(bc\). Players \(j\) and \(k\) in \(G'\) receive the same payoff since in the extensive form they are agents of the same player. It can also be seen that any pair \(q'_k, q'_j\) of strategies for the new players \(j\) and \(k\) has the same effect as the strategy \(q_j\) defined by (2.45) and (2.46).

**Remark:** In the light of the interpretation it is clear that (2.47) and (2.48) also hold for mixed combinations instead of pure ones.
Invariance with respect to player splitting: Let \( L \) be a solution function for a class \( \mathcal{G} \) of normal form games and let \( G = (\Phi, H) \) and \( G' = (\Phi', H') \) be games in \( \mathcal{G} \) such that \( G' \) results from \( G \) by splitting off a player \( k \) at \( \psi_j \). Then we have \( L(G') \rightarrow L(G) \).

Interpretation: Invariance with respect to player splitting requires that \( L(G') \) and \( L(G) \) should prescribe essentially the same behavior. In the case that \( L(G) \) prescribes \( \psi_j \) with probability 1, player \( j \)'s behavior in \( L(G') \) is not restricted by the requirement since it does not matter what he does if player \( k \) selects \( \psi_j \).

Impossibility theorem: Let \( \mathcal{G} \) be a complete class of normal form games which contains all games with at most 4 players and at most 3 strategies for every player. Let \( L \) be a solution function for \( \mathcal{G} \) which satisfies the requirements of cell consistency and truncation consistency and which for 2x2-games with two strong equilibrium points either agrees with the proposed solution function (section 10) or with the pure risk dominance solution function (section 9). Then \( L \) does not satisfy the requirement of invariance with respect to player splitting.

Proof: Assume that \( L \) has all properties mentioned in the theorem, including invariance with respect to player splitting. It will be shown that two different ways of finding the solution of a 3x3-game lead to a contradiction. This game \( G \) is shown on the top of Figure II-21.

Consider the game \( G' \) which results from \( G \) by first splitting off a player 3 at player 1's strategy b and then splitting off a player 4 at player 2's strategy b. The result is found in Figure II-21 if one follows the left arrow leading away from \( G \).

The game \( G' \) has a cell formed by players 1 and 2. It can be seen immediately that for fixed strategies of 3 and 4 the payoffs of players 1 and 2 are non-negative linear transforms of the payoffs obtained in the upper left bimatrix.
Figure II-21: Proof of the impossibility theorem. In $G^1$ and $G^2$ the common payoffs of players 1 and 3 are shown in the upper left corner and the common payoffs of players 2 and 4 are shown in the lower right corner. Players 1 and 2 choose between $a$ and $c$ in $G^1$ and between $b$ and $c$ in $G^2$. Players 3 and 4 choose between $ac$ and $b$ in $G^1$ and between $a$ and $bc$ in $G^2$. 

$L(G^3) = (b,b)$

$L(G^5) = (a,a)$


Figure II-22: Extensive form whose agent normal form agrees with the game $G^1$ in figure II-21
Figure II-23: Extensive form whose agent normal form agrees with the game $G^2$ in figure II-21
The Nash-product criterion immediately shows that \((c,c)\) is the solution of the cell.

Note that the issue of payoff dominance does not arise since there is no payoff dominance between the three strong equilibrium points \((a,a)\), \((b,b)\) and \((c,c)\).

\(G^3\) is the truncation of \(G^1\) with respect to the cell formed by 1 and 2. The Nash-product criterion shows that \((L,L)\) is the solution of this game. It follows that players 3 and 4 both must choose \(b\) in \(L(G^1)\). Obviously, \((b,b)\) corresponds to \(L(G^1)\) in \(G\). Consequently, we must have \(L(G) = (b,b)\).

A similar argument is shown on the right side of figure II-21. The game \(G^2\) results from \(G\) by splitting off players 3 and 4 at the strategies \(a\) of both players. In \(G^1\) player 1 and 2 form a cell whose solution is \((b,b)\). The truncated game \(G^5\) has the solution \(L(G^5) = (a,a)\). Consequently, we must have \(L(G) = (a,a)\). This is a contradiction to \(L(G) = (b,b)\).

Remark: It is interesting to ask the question whether the result could be avoided by a more restrictive definition of a cell which would narrow down the applicability of the cell and truncation consistency requirements. In any case, a more restrictive definition of a cell would have to cover the case of a subgame of an extensive form. In this connection, it is worth pointing out that extensive forms can be found whose agent normal forms agree with \(G^1\) and \(G^2\), where the cells formed by 1 and 2 correspond to subgames. These extensive forms are shown in figures II-22 and II-23. Consequently, the impossibility result cannot be avoided by a more restrictive definition of cells.

Interpretation: We must draw the conclusion that it is by no means irrelevant whether a choice between \(a\), \(b\) and \(c\) has a sequential structure or not. Games where a simultaneous choice has to be made can be different from others where the decision is split into two steps involving choices between \(ab\) and \(c\) and between \(a\) and \(b\). If we do not want to give up the idea of a solution function altogether we must abolish one of the
properties which lead to the impossibility result. Among those properties invariance with respect to player splitting seems to be the least compelling one. Upon reflection it does not appear to be an unreasonable idea that risk comparisons between three alternatives may be changed by the imposition of a sequential structure.

After all, one must think of the fact that after a decision between ab and c has been made, in favor of ab, alternative c has become irrelevant and the risk comparisons may look quite different from those which would arise in a simultaneous choice situation. Different sequential orders may require different ways of looking at the situation. Even if it is not easy to understand why this should be so, it is reasonable to suppose that the basic reason for the impossibility result must be searched in this direction.

The proof of the impossibility theorem makes use of the fact that both ways of sequential player splitting in figure II-21 reduce the risk dominance comparisons to comparisons in 2x2-games which result from G by removing either a or b or c from the strategy sets of both players. The three comparisons which can be made in this way result in an intransitive pattern: (a,a) dominates (b,b) and (b,b) dominates (c,c) but (c,c) dominates (a,a). Moreover, each of both ways of sequential player splitting removes one of the three comparisons, namely that between (a,a) and (b,b) in the case of $G^1$, and that between (c,c) and (a,a) in the case of $G^2$. In this way we can see already here that the impossibility result is connected to intransitivities of risk dominance. We shall return to the phenomenon in chapter 4, section 3, after the introduction of our general definition of risk dominance.
14. Splitting into identical types

In the last section we came to the conclusion that we have to reject invariance with respect to sequential agent splitting as a desirable property. In the following we shall look at another way of substituting two players for one player. This kind of player splitting has a natural interpretation in terms of games of incomplete information. In such games a player may have different types which differ with respect to hidden variables known to the player himself but not to the other players. The information on the other players' types takes the form of a probability distribution over type combinations (Harsanyi).

In the extensive form different types have different information sets. Therefore, the agent normal form will treat them as either different players if each type has only one information set or as non-intersecting groups of players in the more general case.

Suppose that one of the players, say player 1, has two types I and II whereas all other players have only one type. Assume that type I occurs with probability $\alpha$ and type II with probability $1-\alpha$. The other players do not know which of both types will make the decisions of player 1 but they know the probabilities. In general the interest of such game models lies in the fact that different types may have different payoff functions or strategy sets, but for us the possibility that both types I and II are identical in every relevant respect is of special theoretical importance. We may think of a situation where the hidden variable known only to player 1 is his hair colour which does not have any strategic significance for the game. In such cases it should not make any difference whether a type distinction is made or not. The solution should remain essentially the same if a player is split into two identical types.
Splitting a player into identical types: Let $G = (\Phi, H)$ be a game with $\Phi = X \Phi i$ and let $j \in N$ be a player. We construct a game $G' = (\Phi', H')$ whose player set $N' = N \cup \{k\}$ contains an additional player $k \in N$. The pure strategy sets $\Phi'_i$ and the payoff function $H'$ of $G'$ are as follows:

(2.49) $\Phi'_i = \Phi_i$ for $i \in N$

(2.50) $\Phi'_k = \Phi_j$

(2.51) $H'_j(\varphi'_j \varphi'_k \varphi'_{-i}) = H_j(\varphi'_j \varphi'_{-i})$

(2.52) $H'_k(\varphi'_j \varphi'_k \varphi'_{-i}) = H_j(\varphi'_k \varphi'_{-i})$

(2.53) $H'_i(\varphi'_j \varphi'_k \varphi'_{-i}) = \alpha H_i(\varphi'_j \varphi'_{-i}) + (1 - \alpha) H_i(\varphi'_k \varphi'_{-i})$

for every $i \in N \setminus \{j\}$

where $\alpha$ is a number with $0 < \alpha < 1$. This game $G'$ is called the game which results from $G$ by splitting player $j$ into identical types $j$ and $k$ with probabilities $\alpha$ and $1 - \alpha$.

Let $q' = (q'_i)_N$ be a mixed strategy combination for $G'$. The strategy combination $q = (q_i)_N$ for $G$ which corresponds to $q'$ is defined as follows:

(2.54) $q_j(\varphi_j) = \alpha q'_j(\varphi_j) + (1 - \alpha) q'_k(\varphi_j)$

for every $\varphi_j \in \Phi_j$

(2.55) $q'_i = q_i$ for every $i \in N \setminus \{j\}$

It is clear from (2.54) that we have

(2.56) $H_i(q) = H'_i(q')$ for every $i \in N$

if $q$ corresponds to $q'$. We say that $q' = (q'_i)_N$, results from $q = (q_i)_N$ by splitting $j$ into $j$ and $k$ if in addition to (2.55) we have $q'_j = q'_k = q_j$. 


Identical type invariance: Let $L$ be a solution function for a class of games $G$. The solution function $L$ is called invariant with respect to splitting into identical types if the following is true for any two games $G$ and $G'$ in $G$ such that $G'$ results from $G$ by splitting player $j$ into two identical types $j$ and $k$ with probabilities $\alpha$ and $1-\alpha$ where $0 < \alpha < 1$. The solution $L(G')$ of $G'$ results from $L(G)$ by splitting into $j$ and $k$.

**Theorem:** Let $G$ be a complete class of games, let $G_o$ be the subclass of indecomposable games in $G$ and let $L_o$ be the solution function for $G_o$. If $L_o$ is invariant with respect to splitting into identical types, then the extension $L$ of $L_o$ to $G$ is invariant with respect to splitting into identical types.

**Proof:** Let $G'$ be the game which results from $G$ by splitting player $j$ into identical types $j$ and $k$ with probabilities $\alpha$ and $1-\alpha$. Consider a group $C$ of players which forms a cell in $G$. The way in which $H'$ is linearly related to $H$ has the immediate consequence that $C \cup \{k\}$ forms a cell in $G'$. Moreover, it can be seen that $G'$ has no other cells than those which arise in this way. This has the consequence that $G'$ is indecomposable if and only if $G$ is indecomposable. Therefore $L(G')$ results from $L(G)$ by splitting $j$ into $j$ and $k$ if $G$ is indecomposable. On the basis of this fact we prove the theorem by induction on the number $k$ of players in $G$.

Since 1-person games are indecomposable the assertion holds for $n = 1$. Assume that it holds for any number of players up to $n-1$. We have to show that $L(G')$ results from $L'(G)$ by splitting into $j$ and $k$ if $G$ is a decomposable game with $n$ players. Obviously, the elementary cells of $G'$ are either identical to elementary cells of $G$ or they result from such cells by splitting $j$ into two types $j$ and $k$ with probabilities $\alpha$ and $1-\alpha$. The solution of each elementary cell of $G'$ results from that of the corresponding cell of $G$ by splitting $j$ into $j$ and $k$. The assertion follows if $G$ is fully decomposable. If $G$ is not fully decom-
posable then it follows that the main truncation of \( G' \) either is identical to that of \( G \) or it results from it by splitting \( j \) into \( j \) and \( k \) with probabilities \( \alpha \) and \( 1-\alpha \). In the latter case the solution of the main truncation of \( G' \) results from that of \( G \) by splitting into \( j \) and \( k \) since the main truncation of \( G \) has fewer than \( n \) players. It follows that \( L(G') \) results from \( L(G) \) by splitting \( j \) into \( j \) and \( k \).

Comment: As we shall see, the solution function proposed in this book is invariant with respect to splitting into identical types. Certainly, this is a desirable property. As a tool of axiomatic characterization the requirement is probably not a strong one. Nevertheless, it has proved to be quite useful in the search for a reasonable equilibrium point selection theory since it excludes many ideas which would otherwise suggest themselves. It is perhaps sufficient to mention just one example. On the basis of our axiomatic characterization of risk dominance in 2x2-games the following generalization of the definition obtained there suggests itself. Let \( U = (U_i)_N \) and \( V = (V_i)_N \) be two equilibrium points of a game \( G = (\phi,H) \). Define

\[
(2.57) \quad u_i = H_i(U) - H_i(U_iV_{-i})
\]

\[
(2.58) \quad v_i = H_i(V) - H_i(V_iU_{-i})
\]

One is tempted to define risk dominance as follows: \( U \) risk dominates \( V \) if the product \( u_1u_2 \cdots u_n \) is greater than the product \( v_1v_2 \cdots v_n \). This definition fails to be invariant with respect to splitting into identical types since after a splitting of player \( j \) the factors \( u_j \) and \( v_j \) will appear twice in the deviation loss products.
"WIRTSCHAFTSTHEORETISCHE ENTSCHEIDUNGSFORSCHUNG"

A series of books published by the Institute of Mathematical Economics, University of Bielefeld.

Wolfgang Rohde
Ein spieltheoretisches Modell eines Terminmarktes (A Game Theoretical Model of a Futures Market)
The model takes the form of a multistage game with imperfect information and strategic price formation by a specialist. The analysis throws light on theoretically difficult empirical phenomena.

Vol. 1 price: DM 24,80

Klaus Binder
Oligopolistische Preisbildung und Markteintritte (Oligopolistic Pricing and Market Entry)
The book investigates special subgame perfect equilibrium points of a three-stage game model of oligopoly with decisions on entry, on expenditures for market potential and on prices.

Vol. 2 price: DM 22,80

Karin Wagner
Ein Modell der Preisbildung in der Zementindustrie (A Model of Pricing in the Cement Industry)
A location theory model is applied in order to explain observed prices and quantities in the cement industry of the Federal Republic of Germany.

Vol. 3 price: DM 24,80

Rolf Stoecker
Experimentelle Untersuchung des Entscheidungsverhaltens im Bertrand-Oligopol (Experimental Investigation of Decision-Behavior in Bertrand-Oligopoly Games)
The book contains laboratory experiments on repeated supergames with two, three and five bargainers. Special emphasis is put on the end-effect behavior of experimental subjects and the influence of altruism on cooperation.

Vol. 4 price: DM 28,80

Angela Klopstech
Eingeschränkt rationale Marktprozesse (Marketprocesses with Bounded Rationality)
The book investigates two stochastic market models with bounded rationality, one model describes an evolutionary competitive market and the other an adaptive oligopoly market with Markovian interaction.

Vol. 5 price: DM 23,-- appr.

Orders should be sent to:
Pfeffersche Buchhandlung, Alter Markt 7, 4800 Bielefeld 1, West Germany.