Mini-course on Epistemic Game Theory
Lecture 1: Common Belief in Rationality

Andrés Perea
EPICENTER & Dept. of Quantitative Economics

Maastricht University

Singapore, September 2016

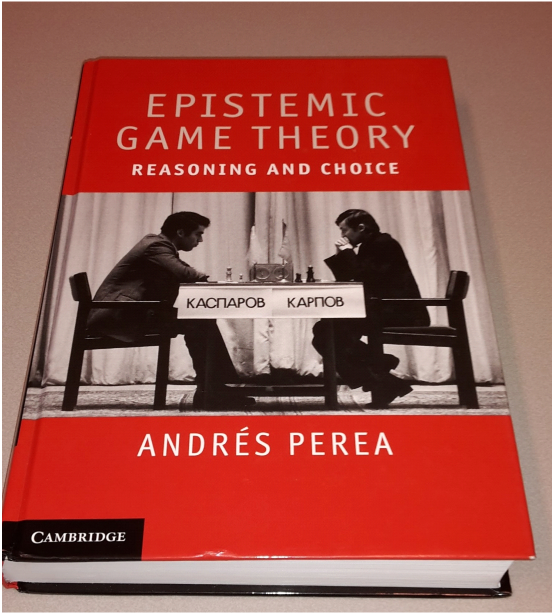
Introduction

- Game theory studies situations where you make a decision, but where the final outcome also depends on the choices of others.
- Before you make a choice, it is natural to reason about your opponents – about their choices but also about their beliefs.
- Oskar Morgenstern, in 1935, already stresses the importance of such reasoning for games.

- Classical game theory has focused mainly on the choices of the players.
- Epistemic game theory asks: Where do these choices come from?
- More precisely, it studies the beliefs that motivate these choices.
- Since the late 80's it has developed a broad spectrum of epistemic concepts for games.
- Some of these characterize existing concepts in classical game theory, others provide new ways of reasoning.

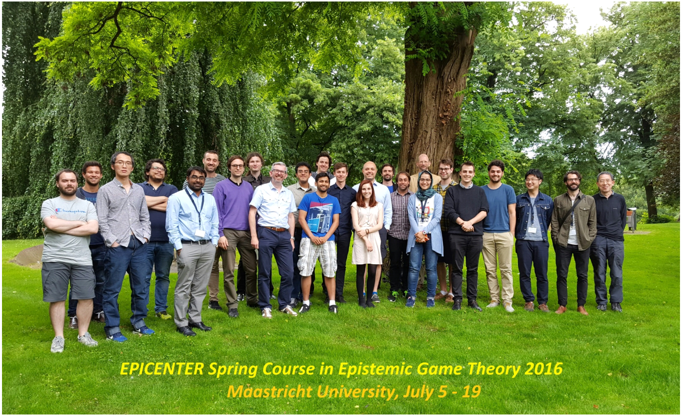
Outline

- In the first lecture, we discuss the idea of common belief in rationality.
- We show that the induced choices are given by iterated elimination of strictly dominated choices.
- In the second lecture we focus on Nash equilibrium.
- We provide an epistemic foundation for Nash equilibrium, and see that it requires more than just common belief in rationality.
- We investigate the extra conditions that lead to Nash equilibrium.
- In the seminar, we will extend these findings to games with incomplete information.


EPISTEMIC
GAME THEORY
REASONING AND CHOICE
КАСПАРОВ
КАРПОВ
ANDRÉS PEREA
CAMBRIDGE

EPICENTER Spring Course in
Epistemic Game Theory

June 12 – 26, 2017
Maastricht University



- If you are an expected utility maximizer, you form a belief about the opponents' choices, and make a choice that is optimal for this belief.
- That is, you choose rationally given your belief.
- It seems reasonable to believe that your opponents will choose rationally as well, ...
- and that your opponents believe that the others will choose rationally as well, and so on.
- Common belief in rationality.

Example: Going to a party

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	2	1	4	3	0

Story

- This evening, you are going to a party together with your friend Barbara.
- You must both decide which color to wear: blue, green, red or yellow.
- Your preferences for wearing these colors are as in the table. These numbers are called utilities.
- You dislike wearing the same color as Barbara: If you both would wear the same color, your utility would be 0.
- What color would you choose, and why?

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	2	1	4	3	0

- Choosing blue is optimal if you believe that Barbara chooses green.
- Choosing green is optimal if you believe that Barbara chooses blue.
- Choosing red is optimal if you believe that, with probability 0.6, Barbara chooses blue, and that with probability 0.4 she chooses green.
- Hence, blue, green and red are rational choices for you.

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	2	1	4	3	0

- Choosing yellow can never be optimal for you, even if you hold a probabilistic belief about Barbara's choice.
- If you assign probability less than 0.5 to Barbara's choice blue, then by choosing blue yourself, your expected utility will be at least $(0.5) \cdot 4 = 2$.
- If you assign probability at least 0.5 to Barbara's choice blue, then by choosing green yourself your expected utility will be at least $(0.5) \cdot 3 = 1.5$.
- Hence, whatever your belief about Barbara, you can always guarantee an expected utility of at least 1.5.
- So, yellow can never be optimal for you, and is therefore an irrational choice for you.

	blue	green	red	yellow	same color as friend
you	4	3	2	×	0
Barbara	2	1	4	3	0

- If you believe that Barbara chooses rationally, and believe that Barbara believes that you choose rationally,

  then you believe that Barbara will not choose blue or green.

	blue	green	red	yellow	same color as friend
you	4	3	2	×	0
Barbara	×	×	4	3	0

- But then, your unique optimal choice is blue.

- So, under common belief in rationality, you can only rationally wear blue.

New Scenario

- Barbara has same preferences over colors as you.
- Barbara likes to wear the same color as you, whereas you dislike this.
-

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

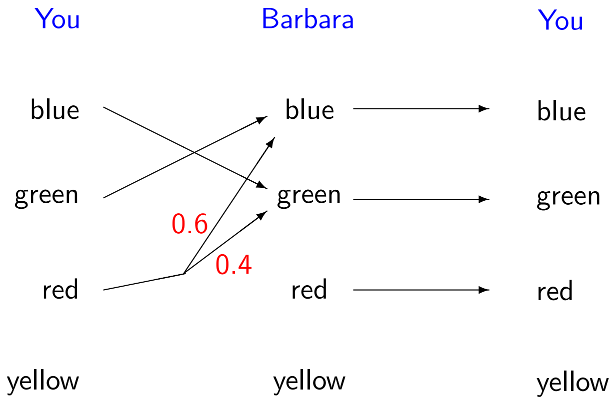
- Which color(s) can you rationally choose under common belief in rationality?

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

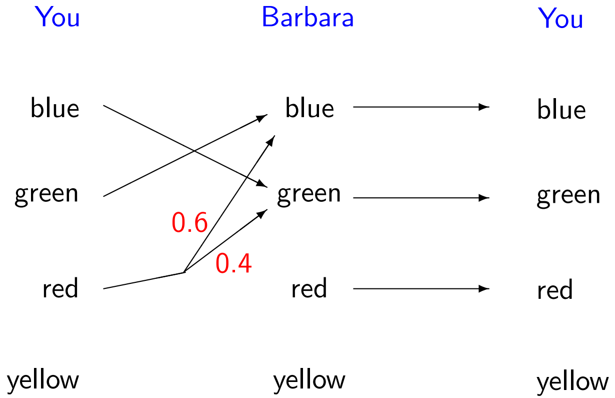
- If you choose rationally, you will not choose yellow.
- If you believe that Barbara chooses rationally, and believe that Barbara believes that you choose rationally, then you believe that Barbara will not choose yellow either.
-

	blue	green	red	yellow	same color as friend
you	4	3	2	×	0
Barbara	4	3	2	×	5

Beliefs diagram



	blue	green	red	yellow	same color as friend
you	4	3	2	×	0
Barbara	4	3	2	×	5



- The belief hierarchy that starts at your choice blue expresses common belief in rationality.
- Similarly, the belief hierarchies that start at your choices green and red also express common belief in rationality.
- So, you can rationally choose blue, green and red under common belief in rationality.

Epistemic model

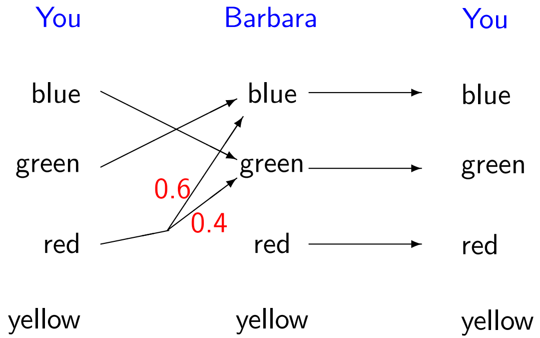
- Writing down a belief hierarchy explicitly is impossible. You must write down
- your belief about the opponents' choices
- your belief about what your opponents believe about their opponents' choices,
- a belief about what the opponents believe that their opponents believe about the other players' choices,
- and so on, ad infinitum.
- Is there an easy way to encode a belief hierarchy?

- A belief hierarchy for you consists of a first-order belief, a second-order belief, a third-order belief, and so on.

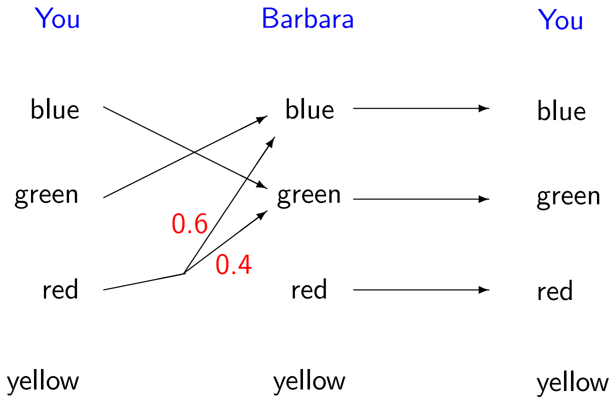
- In a belief hierarchy, you hold a belief about
- the opponents' choices,
- the opponents' first-order beliefs,
- the opponents' second-order beliefs,
- and so on.

- Hence, in a belief hierarchy you hold a belief about
- the opponents' choices, and the opponents' belief hierarchies.

- Following Harsanyi (1967–68), call a belief hierarchy a type.
- Then, a type holds a belief about the opponents' choices and the opponents' types.

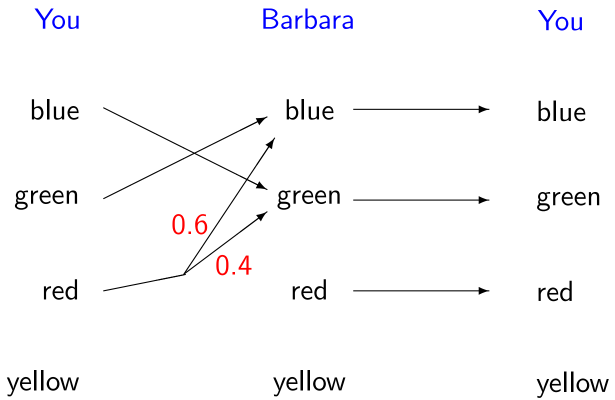


- Denote by t_1^{red} your belief hierarchy that starts at your choice red.
- Denote by t_2^{blue} and t_2^{green} the belief hierarchies for Barbara that start at her choices blue and green.
- Then, t_1^{red} believes that, with prob. 0.6, Barbara chooses blue and has belief hierarchy t_2^{blue}, and believes that, with prob. 0.4, Barbara chooses green and has belief hierarchy t_2^{green}.



- Formally: We call the belief hierarchies t_1^{red}, t_2^{blue} and t_2^{green} types.
- Type t_1^{red} has belief

$$b_1(t_1^{red}) = (0.6) \cdot (blue, t_2^{blue}) + (0.4) \cdot (green, t_2^{green}).$$



- Also, $b_1(t_1^{blue}) = (green, t_2^{green})$ and $b_1(t_1^{green}) = (blue, t_2^{blue})$.
- We can do the same for Barbara's belief hierarchies. This leads to an epistemic model.

Epistemic model for "Going to a party"

Types	$T_1 = \{t_1^{blue}, t_1^{green}, t_1^{red}\}$ $T_2 = \{t_2^{blue}, t_2^{green}, t_2^{red}\}$
Beliefs for player 1	$b_1(t_1^{blue}) = (green,\ t_2^{green})$ $b_1(t_1^{green}) = (blue,\ t_2^{blue})$ $b_1(t_1^{red}) = (0.6) \cdot (blue,\ t_2^{blue}) + (0.4) \cdot (green,\ t_2^{green})$
Beliefs for player 2	$b_2(t_2^{blue}) = (blue,\ t_1^{blue})$ $b_2(t_2^{green}) = (green,\ t_1^{green})$ $b_2(t_2^{red}) = (red,\ t_1^{red})$

- In an epistemic model, we can derive for every type the first-order belief, second-order belief, and so on.
- So, we can derive for every type the complete belief hierarchy .

Types	$T_1 = \{t_1^{blue}, t_1^{green}, t_1^{red}\}$ $T_2 = \{t_2^{blue}, t_2^{green}, t_2^{red}\}$
Beliefs for player 1	$b_1(t_1^{blue}) = (green,\ t_2^{green})$ $b_1(t_1^{green}) = (blue,\ t_2^{blue})$ $b_1(t_1^{red}) = (0.6) \cdot (blue,\ t_2^{blue}) + (0.4) \cdot (green,\ t_2^{green})$
Beliefs for player 2	$b_2(t_2^{blue}) = (blue,\ t_1^{blue})$ $b_2(t_2^{green}) = (green,\ t_1^{green})$ $b_2(t_2^{red}) = (red,\ t_1^{red})$

- Let $I = \{1, ..., n\}$ be the set of players.
- For every player i, let C_i be the finite set of choices.

Definition (Epistemic model)

A finite epistemic model specifies for every player i a finite set T_i of possible types.

Moreover, for every type t_i it specifies a probabilistic belief $b_i(t_i)$ over the set $C_{-i} \times T_{-i}$ of opponents' choice-type combinations.

- Implicit epistemic model: For every type, we can derive the belief hierarchy induced by it.
- This is the model as used by Tan and Werlang (1988).
- Builds upon work by Harsanyi (1967 / 1968), Armbruster and Böge (1979), Böge and Eisele (1979), and Bernheim (1984).

- Remember: A type t_i holds a probabilistic belief $b_i(t_i)$ over the set $C_{-i} \times T_{-i}$ of opponents' choice-type combinations.
- For a choice c_i, let

$$u_i(c_i, t_i) := \sum_{(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}} b_i(t_i)(c_{-i}, t_{-i}) \cdot u_i(c_i, c_{-i})$$

  be the expected utility that type t_i obtains by choosing c_i.
- Choice c_i is optimal for type t_i if

$$u_i(c_i, t_i) \geq u_i(c_i', t_i) \text{ for all } c_i' \in C_i.$$

Definition (Belief in the opponents' rationality)

Type t_i believes in the opponents' rationality if his belief $b_i(t_i)$ only assigns positive probability to opponents' choice-type pairs (c_j, t_j) where choice c_j is optimal for type t_j.

Definition (Common belief in rationality)

(Induction start) Type t_i expresses 1-fold belief in rationality if t_i believes in the opponents' rationality.

(Inductive step) For every $k \geq 2$, type t_i expresses k-fold belief in rationality if t_i only assigns positive probability to opponents' types that express $(k-1)$-fold belief in rationality.

Type t_i expresses common belief in rationality if t_i expresses k-fold belief in rationality for all k.

- Based on Tan and Werlang (1988) and Brandenburger and Dekel (1987).

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

Beliefs for player 1	$b_1(t_1^{blue}) = (green,\ t_2^{green})$ $b_1(t_1^{green}) = (blue,\ t_2^{blue})$ $b_1(t_1^{red}) = (0.6)\cdot(blue,\ t_2^{blue}) + (0.4)\cdot(green,\ t_2^{green})$
Beliefs for player 2	$b_2(t_2^{blue}) = (blue,\ t_1^{blue})$ $b_2(t_2^{green}) = (green,\ t_1^{green})$ $b_2(t_2^{red}) = (red,\ t_1^{red})$

- Each of your types t_1^{blue}, t_1^{green} and t_1^{red} expresses common belief in rationality.
- So, you can rationally choose blue, green and red under common belief in rationality.

Algorithm

- Suppose we wish to find those choices you can rationally make under common belief in rationality.
- Is there an algorithm that helps us find these choices?

- We start with a more basic question: Which choices can be optimal for some belief about the opponents' choices?

Lemma (Pearce (1984))

A choice c_i is optimal for some probabilistic belief about the opponents' choices, if and only if, c_i is not strictly dominated by any randomized choice.

- Here, a randomized choice r_i for player i is a probability distribution on i's choices.

- Choice c_i is strictly dominated by the randomized choice r_i if

$$u_i(c_i, c_{-i}) < u_i(r_i, c_{-i})$$

for every opponents' choice-combination $c_{-i} \in C_{-i}$.

Step 1: 1-fold belief in rationality

- Which choices are rational for a type that expresses 1-fold belief in rationality?
- If you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are optimal for some probabilistic belief.
- Remember: A choice is optimal for some probabilistic belief, precisely when it is not strictly dominated.
- So, if you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are not strictly dominated.

Step 1: 1-fold belief in rationality

- So, if you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are not strictly dominated.
- In a sense, you eliminate the opponents' strictly dominated choices from the game, and concentrate on the reduced game that remains.
- The choices that you can rationally make if you believe in your opponents' rationality, are exactly the choices that are optimal for you for some belief within this reduced game.
- But these are exactly the choices that are not strictly dominated for you within this reduced game.
- Hence, these are the choices that survive 2-fold elimination of strictly dominated choices.

Step 2: Up to 2-fold belief in rationality

- Which choices are rational for a type that expresses up to 2-fold belief in rationality?

- Consider a type t_i that expresses up to 2-fold belief in rationality. Then, t_i only assigns positive probability to opponents' choice-type pairs (c_j, t_j) where c_j is optimal for t_j, and t_j expresses 1-fold belief in rationality.

- So, type t_i only assigns positive probability to opponents' choices c_j which are optimal for a type that expresses 1-fold belief in rationality.

- Hence, type t_i only assigns positive probability to opponents' choices c_j which survive 2-fold elimination of strictly dominated choices.

Step 2: Up to 2-fold belief in rationality

- Hence, type t_i only assigns positive probability to opponents' choices c_j which survive 2-fold elimination of strictly dominated choices.
- Then, every choice c_i which is optimal for t_i must be optimal for some belief within the reduced game obtained after 2-fold elimination of strictly dominated choices.
- So, every choice c_i which is optimal for t_i must not be strictly dominated within the reduced game obtained after 2-fold elimination of strictly dominated choices.
- Conclusion: Every choice that is optimal for a type that expresses up to 2-fold belief in rationality, must survive 3-fold elimination of strictly dominated choices.

Algorithm (Iterated elimination of strictly dominated choices)

Consider a finite static game Γ.

(Induction start) Let $\Gamma^0 := \Gamma$ *be the original game.*

(Inductive step) For every $k \geq 1$, *let* Γ^k *be the game which results if we eliminate from* Γ^{k-1} *all choices that are strictly dominated within* Γ^{k-1}.

- This algorithm terminates within finitely many steps. That is, there is some K with $\Gamma^{K+1} = \Gamma^K$.
- The choices in Γ^K are said to survive iterated elimination of strictly dominated choices.
- It always yields a nonempty set of choices for all players.
- The final output does not depend on the order by which we eliminate choices.

Algorithm (Iterated elimination of strictly dominated choices)

Consider a finite static game Γ.

(Induction start) Let $\Gamma^0 := \Gamma$ *be the original game.*

(Inductive step) For every $k \geq 1$*, let* Γ^k *be the game which results if we eliminate from* Γ^{k-1} *all choices that are strictly dominated within* Γ^{k-1}.

- In two-player games, it yields exactly the rationalizable choices, as defined by Bernheim (1984) and Pearce (1984).
- For games with more than two players, rationalizability requires player i's belief about player j's choice to be stochastically independent from his belief about player k's choice.
- The algorithm does not impose this independence condition.

Theorem (Tan and Werlang (1988))

(1) For every $k \geq 1$, the choices that are optimal for a type that expresses up to k-fold belief in rationality are exactly those choices that survive $(k+1)$-fold elimination of strictly dominated choices.

(2) The choices that are optimal for a type that expresses common belief in rationality are exactly those choices that survive iterated elimination of strictly dominated choices.

- Proof of part (2):
- We already know: If choice c_i is optimal for a type t_i that expresses common belief in rationality, then c_i must survive iterated elimination of strictly dominated choices.

- We now show the converse: If a choice survives iterated elimination of strictly dominated choices, then it can rationally be made under common belief in rationality.
- Assume two players. Suppose that the algorithm terminates after K steps. Let C_i^K be the set of surviving choices for player i.
- Then, every choice in C_i^K is not strictly dominated within the reduced game Γ^K. Hence, every choice c_i in C_i^K is optimal for some belief $b_i^{c_i} \in \Delta(C_j^K)$.
- Define set of types $T_i = \{t_i^{c_i} : c_i \in C_i^K\}$ for both players i.
- Every type $t_i^{c_i}$ only deems possible opponent's choice-type pairs $(c_j, t_j^{c_j})$, with $c_j \in C_j^K$, and

$$b_i(t_i^{c_i})(c_j, t_j^{c_j}) := b_i^{c_i}(c_j).$$

- Then, every type $t_i^{c_i}$ believes in the opponents' rationality.
- Hence, every type expresses common belief in rationality. ∎

Corollary (Common belief in rationality is always possible)

We can always construct an epistemic model in which all types express common belief in rationality.

Example: Guessing two-thirds of the average

Story

- All people in this room must write a number on a piece of paper, between 1 and 100.
- The closer you are to two-thirds of the average of all numbers, the higher your prize money.

- What number(s) could you have rationally written down under common belief in rationality?
- Apply the algorithm of "iterated elimination of strictly dominated choices".
- Step 1: What numbers are strictly dominated?
- Two-thirds of the average can never be above 67.
- Hence, every number above 67 is strictly dominated by 67.
- Eliminate all numbers above 67.

- Step 2: Consider the reduced game Γ^1 in which only the numbers $1, ..., 67$ remain for all people.
- Which numbers are strictly dominated in Γ^1 ?
- Two-thirds of the average of all numbers in Γ^1 can never be above $\frac{2}{3} \cdot 67 \approx 45$.
- All numbers above 45 are strictly dominated in Γ^1.
- Eliminate all numbers above 45.
- And so on.
- Only the number 1 remains at the end.
- Under common belief in rationality, you must choose number 1.
- Would you really choose this number? Why?

- W. Armbruster and W. Böge, 'Bayesian game theory ' in O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Related Topics* (North-Holland, Amsterdam, 1979)
- B.D. Bernheim, 'Rationalizable strategic behavior', *Econometrica,* 52 (1984), 1007–1028
- W. Böge and T.H. Eisele, 'On solutions of bayesian games', *International Journal of Game Theory,* 8 (1979), 193–215
- A. Brandenburger and E. Dekel, 'Rationalizability and correlated equilibria', *Econometrica,* 55 (1987), 1391–1402
- J.C. Harsanyi, 'Games with incomplete information played by "bayesian" players, I–III', *Management Science,* 14 (1967–1968), 159–182, 320–334, 486–502

- O. Morgenstern, ‘Vollkommene Voraussicht und wirtschaftliches Gleichgewicht’, *Zeitschrift für Nationalökonomie,* 6 (1935), 337–357. (Reprinted as ‘Perfect foresight and economic equilibrium’ in A. Schotter (ed.), *Selected Economic Writings of Oskar Morgenstern* (New York University Press, 1976), pp. 169–183)
- D. Pearce, ‘Rationalizable strategic behavior and the problem of perfection’, *Econometrica,* 52 (1984), 1029–1050
- T. Tan and S.R.C. Werlang, ‘The bayesian foundations of solution concepts of games’, *Journal of Economic Theory,* 45 (1988), 370–391