Teams: Building, Adaptation and Learning

CHAPTER · JANUARY 2016

DOI: 10.1007/978-3-319-23338-3_1

1 AUTHOR:

Dmitry A. Novikov
Institute of Control Sciences

New Frontiers in Information and Production Systems Modeling and Analysis – Incentive Mechanisms, Competence Management, Knowledge-based Production / Series "Intelligent Systems Reference Library". Ed. by P. Rozewsky, D. Novikov, O. Zaikin, N. Bakhtadze. – Berlin: Springer, 2016. – 268 p. - P. 3 - 34.

1. TEAMS: BUILDING, ADAPTATION AND LEARNING

Novikov D.A.
(Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, *novikov@ipu.ru*, +74953347569, Profsojuznaya st., 65, Moscow, 117997, Russia)

Game-theoretical models of of team building, team adaptation and team learning are considered for multi-agent organizational systems. It's shown that models of team building and functioning described in terms of reflexive games reproduce the autonomy and coordination of team activity. Team adaptation is considered as the beliefs' updating under the absence of common knowledge among the agents. In the framework of the joint learning model the optimal learning problem is stated and solved as the allocation of the volumes of works performed by agents in certain time intervals.

1.1. Team building

This section focuses on the model of *team building*, where uncertain parameters are the efficiency levels of agents' activity (see also the basic books and papers [9, 11, 12, 19]).

A *team* is a collective (a union of people performing a joint activity and possessing common interests), being able to achieve a goal in an *autonomous* and *self-coordinated* way under the minimum control actions [15, 18, 22].

The following couple of aspects are essential in the definition of a team. The first aspect concerns *goal* achievement, i.e., the final result of a *joint activity* represents a unifying factor for any team. The second aspect is related to the autonomy and self-coordination of team activity; notably, each member of a team shows the behavior required under specific conditions (leading to the posed goal), i.e., the behavior expected by the rest team members [8, 19].

Notwithstanding numerous qualitative discussions in scientific literature, today one would hardly find formal models of team building with non-trivial *mutual beliefs*. Thus, below we describe the model of team building based on the hierarchies of agents' mutual beliefs [16] about the efficiency of their individual activity.

According to existing beliefs, each agent can forecast the actions to-be-chosen by other agents, the individual "costs" to-be-incurred by them and the total costs of the team. Assume that the actions are chosen repeatedly and the reality observed by a certain agent differs from its beliefs. Then the agent has to correct the beliefs and use "new" ones in its choice.

The analysis of *informational equilibria* [16, 22] shows the following. It is reasonable to consider a team as a set of agents whose choices agree with the hierarchy of their mutual beliefs about each other. As a matter of fact, such definition of a team turns out close to the stability and coordination of *informational control* (both require that the real actions or payoffs of agents coincide with the expected ones).

Moreover, an interesting conclusion is that the stability of a team and self-coordination of its functioning can be ensured under the false beliefs of agents about each other. To disturb a false equilibrium, one should report additional information on agents to them.

Consider the set of *agents* $N = \{1, 2, \dots, n\}$. The strategy of agent i is choosing an *action* $y_i \geq 0$, which requires the costs $c_i(y_i, r_i)$. Here $r_i > 0$ means the type of the agent, reflecting the efficiency of its activity.[1] In this section, we study the Cobb-Douglas cost functions $c_i(y_i, r_i) = y_i^{\alpha} r_i^{1-\alpha}/\alpha$, $i \in N$. Let the goal of agents' joint activity [3] lie in achieving the total "action"

[1] By assumption, the costs of an agent represent a decreasing function of its type.

$$\sum_{i\in N} y_i = R \qquad (1.1)$$

under the minimum total costs:

$$\sum_{i\in N} c_i(y_i, r_i) \to \min_{\{y_i\}\geq 0}. \qquad (1.2)$$

From the game-theoretic viewpoint, for the sake of convenience one may believe that the goal functions of agents coincide and equal the total costs with negative sign.

In practice, this problem possesses the following interpretations: order execution by a production association, performing a given volume of work by a team (a department), and so on. Without loss of generality, set $R = 1$.

Suppose that the vector $r = (r_1, r_2, \ldots, r_n)$ is *common knowledge* [2, 16]. By solving the constrained optimization problem (1.1)-(1.2), each agent evaluates the optimal action vector

$$y^*(r) = (y_1^*(r),\ y_2^*(r),\ \ldots,\ y_n^*(r)),$$

where

$$y_i^*(r) = r_i / \sum_{j\in N} r_j,\ i \in N. \qquad (1.3)$$

Now, discuss different variants of agents' awareness about the type vector (the model considered includes an hierarchy of agents' beliefs about the parameters of each other) [16, 21]. We will explore two cases as follows. In the first case, each agent has certain first-order beliefs $r_{ij} > 0$ about the types of other agents; in the second one, each agent has certain second-order beliefs $r_{ijk} > 0$ about the types of other agents, $i, j, k \in N$.

Making a digression, note that there may exist a *Principal* who knows the true types of agents and performs motivational control (see Chapter 1.). Consequently, regardless of agents' awareness, each agent would independently choose the corresponding action (1.3), if the Principal adopts a proportional incentive scheme with the wage rate $1/\sum_{j\in N} r_j$ [22].

Suppose that each agent knows its type. Moreover, the *axiom of self-awareness* [16] implies that

$$r_{ii} = r_i,\ r_{iij} = r_{ij},\ r_{ijj} = r_{ij},\ i, j \in N.$$

According to its beliefs, each agent can forecast the actions to-be-chosen by other agents, the individual costs to-be-incurred by them, and the resulting total costs. Assume that actions are chosen repeatedly and the reality observed by an agent differs from its beliefs. Thus, an agent should correct the beliefs and involve "new" beliefs for the next choice.

The set of parameters observed by agent i is said to be its *subjective game history* and denoted by h_i, $i \in N$. Within the framework of the present model, the subjective game history of agent i may include the following information:

1) the actions chosen by other agents (of course, agent i is aware of its own choice):

$$y_{-i} = (y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n);$$

2) the actual costs of other agents (and their total costs):

$$c_{-i} = (c_1, c_2, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n);$$

3) the total costs of all agents: $c = \sum_{i\in N} c_i$;

4) the actions and actual costs of other agents (and their total costs): $(y_{-i}; c_{-i})$;

5) the actions of other agents and the total costs of all agents: $(y_{-i}; c)$.

Apparently, the stated variants are not equivalent. Variant 4 seems the most "informative," variant 3 is less informative in comparison with variant 2, etc. The choice of an awareness variant is a tool of *informational control* [16, 22] applied by a Principal.

The two cases of awareness structures (beliefs in the form r_{ij} and in the form r_{ijk}) and the five variants of the subjective game history induce Models 1-10 combined in Table 1.1. We believe that the subjective histories and awareness structures of all agents coincide (otherwise, the number of possible combinations grows exponentially).

Table 1.1. *The models of team building*

Subjective game history	Awareness structure	
	$\{r_{ij}\}$	$\{r_{ijk}\}$
y_{-i}	Model 1	Model 6
c_{-i}	Model 2	Model 7
c	Model 3	Model 8
$(y_{-i}; c_{-i})$	Model 4	Model 9
$(y_{-i}; c)$	Model 5	Model 10

To proceed, we analyze possible decision-making procedures adopted by agents. Under the awareness structure $\{r_{ij}\}$, agent i can choose its action according to the procedure (1.3):

$$y_i^*(\{r_{ij}\}) = r_i / \sum_{j \in N} r_{ij} . \qquad (1.4)$$

Alternatively, the agent may first estimate the opponents' actions via the procedure (2), then evaluate its action leading to the required sum of actions:

$$y_i^*(\{r_{ij}\}) = 1 - \sum_{k \neq i} (r_{ik} / \sum_{l \in N} r_{il}),\ i \in N. \qquad (1.5)$$

Evidently, the procedures (1.4) and (1.5) are equivalent.

Under the awareness structure $\{r_{ijk}\}$, agent i can estimate the opponents' actions using the procedure (1.3):

$$y_{ij}^*(\{r_{ijk}\}) = r_{ij} / \sum_{l \in N} r_{ijl} ,\ j \in N. \qquad (1.6)$$

Then the agent evaluates its action ensuring the required sum of actions:

$$y_i^*(\{r_{ijk}\}) = 1 - \sum_{l \neq i} (r_{il} / \sum_{q \in N} r_{ijq}),\ i \in N. \qquad (1.7)$$

Thus, we have described the models of agents' decision-making in the static mode. Now, consider the *dynamics of their collective behavior*.

Suppose that at each step agents make their decisions simultaneously using the information on the previous step only. In other words, the subjective game history includes the corresponding values during the previous period merely. Such assumption eliminates the case when decision-making takes place on the basis of the whole preceding game trajectory observed by an agent. The underlying reason is that these models of decision-making appear extremely complicated and would hardly yield practically relevant conclusions.

Denote by $W_i^t(h_i^t)$ the current *state of the goal* of agent i during the time period (step) t, i.e., its beliefs I_i^t about the opponents' types that could lead to their choices observed by the agent during this period; here $t = 0, 1, \dots$ and $i \in N$.

Initially, agents possess the beliefs I_i^t and modify them depending on the subjective game history according to the *hypothesis of indicator behavior* [16, 22]:

$$I_i^{t+1} = I_i^t + \gamma_i^t (W_i^t(h_i^t) - I_i^t),\ t = 0, 1, \dots,\ i \in N, \qquad (1.8)$$

where γ_i^t means a vector of real values from the segment [0; 1] ("step sizes").

The beliefs of each agent are described by a finite number of the parameters r_{ij} or r_{ijk}, $i, j, k \in N$. Hence, we will understand (1.8) as the "vector"-form law of independent variations in awareness structure components.

Now, we suggest a rigorous definition of a team. Notably, a *team* is a set of agents whose choices agree with the hierarchy of their mutual beliefs. In the model under consideration, a team is a set of agents with the awareness structure representing a fixed point of the mapping (1.8) provided that the actions chosen by agents according to their awareness structures are given by formula (1.5) or (1.7). The introduced definition of a team is close to the properties of *stability* and *coordination of informational control*; the latter require that the real actions or payoffs of agents coincide with the expected actions or payoffs (see [5, 6, 7, 16]).

Therefore, in each case the dynamics of agents' mutual beliefs is described by the relationship $W_i^t(\cdot)$ between the current state of the agent's goal and its subjective game history.

Model 1. Suppose that agent i with the awareness structure $\{r_{ij}\}$ observes the actions x_{-i} chosen by the opponents.

For agent i, take the set of opponents' types such that their actions chosen according to formula (1.4) coincide with the observed actions x_{-i}. Denote this set by

$$\Omega_i^1 = \{r_{ij} > 0, j \in N \setminus \{i\} \mid r_{ij} / \sum_{j\in N} r_{ij} = x_j, j \in N \setminus \{i\}\}. \quad (1.9)$$

Next, let $w_{ij}^t(x_{-i}^t)$ be the jth projection of the point belonging to the set Ω_i^1 and being closest to the point $(r_{ij}^t)_{j \in N \setminus \{i\}}$. Then the dynamics of the beliefs of agent i can be rewritten as

$$r_{ij}^{t+1} = r_{ij}^t + \gamma_{ij}^t\,(w_{ij}^t(x_{-i}^t) - r_{ij}^t),\ j \in N\setminus\{i\},\ t = 0, 1, \dots,\ i \in N. \quad (1.10)$$

And its choice of actions would follow the expression (2).

Model 2. Assume that agent i with the awareness structure $\{r_{ij}\}$ observes the costs c_{-i} of other agents.

For agent i, take the set of opponents' types such that their costs incurred by the choice of actions according to formula (1.4) coincide with the observed costs c_{-i}. Denote this set by

$$\Omega_i^2 = \{r_{ij} > 0, j \in N \setminus \{i\} \mid c_j(r_{ij} / \sum_{l\in N} r_{il}, r_{ij}) = c_j, j \in N\setminus\{i\}\}. \quad (1.11)$$

Again, let $w_{ij}^t(c_{-i})$ be the jth projection of the point belonging to the set Ω_i^2 and being closest to the point $(r_{ij}^t)_{j \in N \setminus \{i\}}$. Then the dynamics of the beliefs of agent i can be characterized by the procedure (1.10), and its choice of actions agrees with (1.4).

In the sense of informativeness and the feasibility of the corresponding equations (see formulas (1.9) and (1.11)), this case slightly differs from Model 1.

Model 3. Assume that agent i with the awareness structure $\{r_{ij}\}$ observes the costs c of all agents.

For agent i, take the set of opponents' types such that their total costs incurred coincide with the observed total costs c. Denote this set by

$$\Omega_i^3 = \{r_{ij} > 0, j \in N\setminus\{i\} \mid c_i(y_i, r_i) + \sum_{l\in N\setminus\{i\}} [c_j(r_{il} / \sum_{q\in N} r_{iq}, r_{ij})] = c\}. \quad (1.12)$$

Similarly, let $w_{ij}^t(c)$ be the jth projection of the point belonging to the set Ω_i^3 and being closest to the point $(r_{ij}^t)_{j \in N \setminus \{i\}}$. Then the dynamics of the beliefs of agent i can be modeled by the procedure (8), and its choice of actions agrees with (2).

As a matter of fact, this case substantially varies from Models 1–2 (in the aspects of informativeness and non-unique solutions to the equation defining the set Ω_i^3 (see formula (1.12)), as well as in the aspect of modeling complexity).

Models 4–5 are treated by analogy to Models 1–2; thus, their detailed treatment is omitted.

Model 6. Assume that agent i with the awareness structure $\{r_{ijk}\}$ observes the actions x_{-i} chosen by the opponents.

For agent i, take the set of opponents' types such that their actions being chosen by the procedure (1.6) coincide with the observed actions x_{-i}. Denote this set by

$$\Omega_i^6 = \{r_{ijk} > 0, j \in N \setminus \{i\}, k \in N \mid r_{ij} / \sum_{l\in N} r_{ijl} = x_j, j \in N\setminus\{i\}\}. \quad (1.13)$$

Moreover, let $w_{ijk}^t(x_{-i}^t)$ be the jkth projection of the point belonging to the set Ω_i^6 and being closest to the point $(r_{ijk}^t)_{j \in N \setminus \{i\}}$. Then the dynamics of the beliefs of agent i can be defined by

$$r_{ijk}^{t+1} = r_{ijk}^t + \gamma_{ij}^t\,(w_{ijk}^t(x_{-i}^t) - r_{ijk}^t),\ j \in N \setminus \{i\},\ t = 0, 1, \dots,\ i \in N. \quad (1.14)$$

and its choice of actions satisfies the expression (1.7), i.e.,

$$y_i^{t^*}(\{r_{ijk}^t\}) = 1 - \sum_{j\neq i} (r_{il}^t / \sum_{q\in N} r_{ijq}^t)\ i \in N. \quad (1.15)$$

Model 6 appears equivalent to Model 1 in description and analysis techniques. On the other hand, Model 7 is equivalent to Model 2, and so on. Therefore, Models 7–10 are not studied here.

Thus, for each agent Model 1 includes $(n-1)$ equations with $(n-1)$ unknown quantities, Model 2 includes $(n-1)$ equations with $(n-1)$ unknown quantities, Model 3 includes 1 equation with $(n-1)$ unknown quantities, Model 4 includes $2(n-1)$ equations with $(n-1)$ unknown quantities, Model 5 includes n equations with $(n-1)$ unknown quantities, Model 6 includes $(n-1)$ equations with $n(n-1)$ unknown quantities, and so on.

Concluding this section, we consider the simplest model, *viz*., Model 1, in the case of three agents with the separable quadratic cost functions $c_i(y_i, r_i) = (y_i)^2 / 2\, r_i$.

***Model 1* (*an example*).** It follows from formula (1.9) that

$$w_{13}(x_2, x_3) = x_3\, r_1 / (1 - x_2 - x_3),\; w_{12}(x_2, x_3) = x_2\, r_1 / (1 - x_2 - x_3),$$
$$w_{21}(x_1, x_3) = x_1\, r_2 / (1 - x_1 - x_3),\; w_{23}(x_1, x_3) = x_3\, r_2 / (1 - x_1 - x_3),$$
$$w_{31}(x_1, x_2) = x_1\, r_3 / (1 - x_1 - x_2),\; w_{32}(x_1, x_2) = x_2\, r_3 / (1 - x_1 - x_2).$$

Set $r_1 = 1.8$, $r_2 = 2$, $r_3 = 2.2$, and take the initial agents' beliefs about their types equal to 2. The objectively optimal action vector (in the sense of the minimum total costs) makes up (0.30; 0.33; 0.37).

Suppose that agents act in the following way. Based on their own beliefs about their types and the types of the opponents, agents evaluate the opponents' actions guaranteeing the "subjective" total minimum to the costs sum (i.e., forecast the actions of the opponents). This is done according to the procedure (1.4). Next, agents compare the observed actions with the forecasted ones and modify their beliefs about the opponents' types proportionally to the difference between the observed and forecasted actions; the proportionality coefficient constitutes $\gamma_{ij}^{t} = 0.25$, $i, j \in N$, $t = 0, 1, \ldots$

After 200 steps the stated procedure converges to the action vector (0.316; 0.339; 0.345) and the following beliefs of the agents about their types: $r_{12} = 1.93 < r_2$, $r_{13} = 1.94 < r_3$, $r_{21} = 1.86 > r_1$, $r_{23} = 2.01 < r_3$, $r_{31} = 2.02 > r_1$, and $r_{32} = 2.17 > r_2$. Despite the evident mismatches between the reality and the existing beliefs of agents, the outcome appears stable–the expected actions and the observed ones coincide to four digits after the decimal point.

Now, set $r_1 = 1.8$, $r_2 = 2$, $r_3 = 2.2$ and choose other initial beliefs of the agents about their types:

$$r_{12}^{0} = 2,\; r_{13}^{0} = 2.5,\; r_{21}^{0} = 1.5,\; r_{23}^{0} = 2.5,\; r_{31}^{0} = 1.5,\; r_{32}^{0} = 2.$$

Still, the action vector (0.30; 0.33; 0.37) is objectively optimal (in the sense of the minimum total costs).

After 200 steps, the procedure brings to the action vector (0.298; 0.3484; 0.3524) and the following beliefs of the agents about their types: $r_{12} = 2.1 > r_2$, $r_{13} = 2.12 < r_3$, $r_{21} = 1.71 < r_1$, $r_{23} = 2.01 < r_3$, $r_{31} = 1.85 > r_1$, and $r_{32} = 2.16 > r_2$. Again, despite the evident mismatches between the reality and the existing beliefs of agents, the outcome appears stable–the expected actions and the observed ones coincide to four digits after the decimal point.

Under the same initial data, the procedure (1.10) leads to the action vector (0.318; 0.341; 0.341) and the following beliefs of the agents about their types: $r_{12} = 1.93 < r_2$, $r_{13} = 1.93 < r_3$, $r_{21} = 1.87 > r_1$, $r_{23} = 2.00 < r_3$, $r_{31} = 1.05 > r_1$, and $r_{32} = 2.2 > r_2$. Similarly, despite the evident mismatches between the reality and the existing beliefs of agents, the outcome appears stable–the expected actions and the observed ones coincide to six digits after the decimal point.

The phenomenon of informational equilibrium instability (when the mutual beliefs of agents do not coincide with the reality) has a simple explanation. The system of equations (1.9) in the beliefs and actions of all agents possesses non-unique solutions. Indeed, in the case of two agents, the system of three equations

$$\begin{cases} \dfrac{r_{12}}{r_1 + r_{12}} = x_2 \\ x_1 + x_2 = 1 \\ \dfrac{r_{21}}{r_2 + r_{21}} = x_1 \end{cases} \qquad (1.16)$$

with four unknown quantities r_{12}, r_{21}, x_1, x_2 admits an infinite set of solutions. Really, by expressing all unknown quantities through x_1, one we obtain the family of solutions $r_{12} = r_1(1/x_1 - 1)$, $r_{21} = r_2\, x_1/(1 - x_1)$, $x_2 = 1 - x_1$, and $x_1 \in (0; 1)$. Substitution of these beliefs into (1.4) yields identities.

Interestingly, transition to Model 4, i.e., adding information on the opponents' costs may considerably reduce the set of solutions for the corresponding system of equations. In this model, simultaneous observation of the costs and actions of an agent enables unique definition of its type (in one step).

We provide an example. Consider two agents with the types $r_1 = 1.5$ and $r_2 = 2.5$. The initial (appreciably "incorrect") beliefs are $r_{12}^0 = 1.8$ and $r_{21}^0 = 2.2$. After 200 steps, the resulting mutual beliefs of the agents make up $r_{12} = 1.747$ and $r_{21} = 2.147$, i.e., still being far from the truth.

At the same time, the subjectively equilibrium actions constitute $x_1 = 0.4614$ and $x_2 = 0.5376$. The actions observed by the agents form an informational equilibrium–they are coordinated with the individual beliefs of the agents (i.e., satisfy the system of equations (1.16)).

For the example above, the set of subjective equilibria is illustrated by Fig. 1.1 (here the circle indicates the initial point, the diamond stands for the actual values of the types, while the arrow shows the direction of variation of the agents' beliefs).

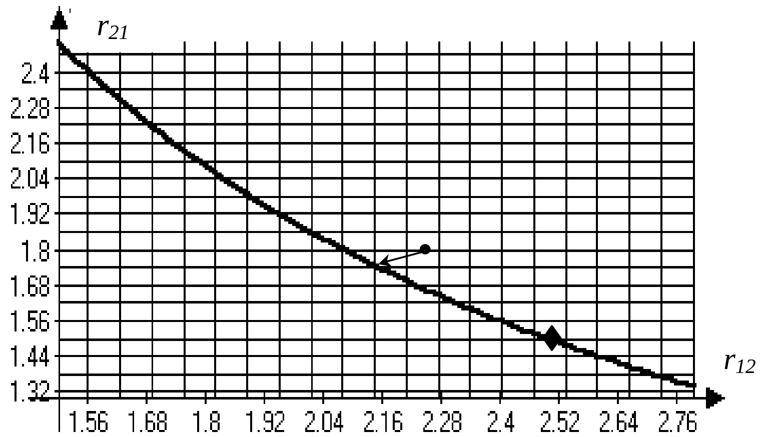

Fig. 1.1. *The set of subjective equilibria*

The system of equations (1.16) implies that stable informational equilibria satisfy the following condition:

$$r_{12}\, r_{21} = r_1\, r_2. \qquad (1.17)$$

The set of mutual beliefs (r_{12}; r_{21}) meeting (1.17) is a hyperbola on the corresponding plane. Fig. 1.2 demonstrates such hyperbola in the case of $r_1 = 2$ and $r_2 = 1$.

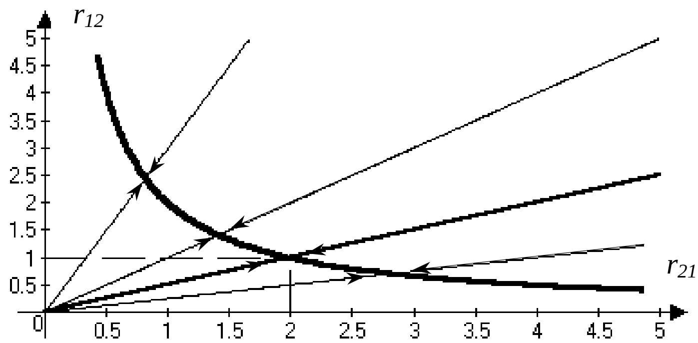

Fig. 1.2. *The set of subjective equilibria and their domains of attraction*

The performed analysis makes it possible to define the set of *false equilibria* (1.17), as well as to study the corresponding domains of attraction. It follows from (1.10) that the dynamics of the mutual beliefs obeys the equation

$$\frac{\Delta r_{12}^{t+1}}{\Delta r_{21}^{t+1}} = \frac{\gamma_{12}^{t+1}}{\gamma_{21}^{t+1}} \; \frac{r_{12}^{t}}{r_{21}^{t}}, \; t = 0, 1, \ldots \qquad (1.18)$$

Hence, under fixed and constant "steps" γ, the trajectories of the mutual beliefs are lines passing through the origin. The slope of these lines (the domains of attraction for the intersection points with the hyperbola (1.17)) depends on the initial point. For instance, any initial point lying on the thick line in Fig. 1.2 ($r_{12} = r_{21} / 2$) leads to the true equilibrium.

This fact seems interesting in the sense of informational control. The Principal can easily evaluate the set of initial points (a line) leading to a desired terminal point (agents would definitely reach the equilibrium[2] required to the Principal). Note that the Principal must have adequate information on the types of agents.

The above example leads to the conclusion that the stability and congruence of a team can be achieved under the false mutual beliefs of team members. Disturbing a false equilibrium requires additional information on agents.

Therefore, the models of team building and functioning described in terms of reflexive games reproduce the autonomy and self-coordination of a team. Furthermore, they enable posing and solving control problems for the process of team building.

Indeed, the study of Models 1–10 shows that the most important information on the game history is the one available to agents. Thus, control capabilities lie in

- creating different situations of activity and
- ensuring maximum communication and access to relevant information.

The analysis also testifies that the *speed of team building* (the rate of convergence to an equilibrium) essentially depends on the parameters $\{\gamma\}$ ("step sizes") used in the dynamic behavioral procedures of agents. The impact on these parameters can be treated as control by the Principal.[3]

A natural question arises immediately. Is the outcome of a false equilibrium frequent? We elucidate this issue by stating the general conditions of its occurrence (within the framework of Model 1).

Suppose that the agents' type vector $r = (r_1, r_2, ..., r_n)$ is common knowledge, and there exists a unique optimal action vector $y^*(r) = (y_1^*(r), y_2^*(r), ..., y_n^*(r))$.

Hence, we have defined n functions $\varphi_i: r \to y_i^*(r)$, $i \in N$, mapping the type vector r into the optimal action of agent i (the domains of the functions φ_i include only type vectors leading to a unique vector of optimal actions).

Now, assume that the outcome described above takes place *subjectively*; i.e., each agent *believes* that the type vector is common knowledge. Then the awareness structure of the game is defined by N vectors of the form $(r_{i1}, r_{i2}, \ldots, r_{in})$, $i \in N$. The informational equilibrium $y^* = (y_1^*, y_2^*, ..., y_n^*)$ is stable if each agent observes the actions of the opponents expected by it. This means that

$$\varphi_i (r_{j1}, r_{j2}, \ldots, r_{jn}) = y_i^*, \; i, j \in N. \qquad (1.19)$$

The equilibrium y^* being arbitrary, formula (17) specifies n^2 constraints on the awareness structure. Next, if the type of each agent is fixed (and each agent knows its type), then one should evaluate $n(n-1)$ quantities r_{ij}, $i, j \in N$, $i \neq j$ in order to guarantee the expressions (1.19).

The system (1.19) is *a fortiori* satisfied by the set of quantities r_{ij} such that $r_{ij} = r_j$ for all i and j. Therefore, under a fixed set of types $(r_1, r_2, \ldots, r_n)$, the issue regarding the existence of a false equilibrium is reduced to the following question. Does the system (1.19) admit non-unique solutions?

One may put the following hypothesis. The outcome of a false equilibrium is rather an exception, and its occurrence in the discussed examples is connected with a specific interaction among agents. This hypothesis is confirmed by some examples in [16] (no false equilibria are found there).

The conducted analysis brings to the following conclusion. The models of team building and functioning described in terms of *reflexive games* reproduce the autonomy and coordination of team activity. Furthermore, they allow posing and solving control problems for team building process. Control capabilities include, first, creating various situations of activity (to identify the essential characteristics of agents–the model of learning) and, second, ensuring maximum communication and access of team members to all essential information.

1.2. Team adaptation

Problems of team adaptation. This section considers the models of *team adaptation*, i.e., the process of changes in the actions chosen by team members based on current information in varying external condi-

---

[2] In the case of variable "steps," the problem is to find a trajectory which satisfies (1.18) and passes through a given point of the set (1.18).

[3] Note that increasing the step size improves the rate of convergence; on the other hand, the procedure may become instable for sufficiently large step sizes.

tions of team functioning [19]. In the general case, adaptation also covers the functions and volumes of work performed by team members. We separate out several embedded *levels of adaptation*, namely:

1) changing of awareness about an external environment;

2) changing of behavior (actions chosen using available information);

3) changing of system parameters, which enables implementing more efficient behavior in varying conditions (*learning*);

4) purposeful modification of an external environment (*active adaptation*).

Particularly, it is demonstrated that *teams possess the following specifics*. Each agent corrects its beliefs about an uncertain parameter based on the observations of an external environment and, moreover, the observations of the actions and results of other agents (endeavoring to "explain" their choice [19]). In other words, if the result of joint activity depends on the actions of all agents, each agent operates (at most) four "sources of information" on the external environment:

1) the a priori individual information possessed by each agent;

2) the actions of other agents (by observing them and assuming the rational decision-making of opponents, an agent may perform reflexion, i.e., assess the information on the external environment leading to exactly these rational choices of the opponents);

3) the payoffs of agents (using this information, agents make conclusions on the states of the external environment, under which the observed result leads to the observed payoffs);

4) the set of states of the external environment, under which the observed vector of agents' actions leads to the observed result).

We introduce the notion of *adaptation time* in a team; this is the time required for the unambiguous identification of the state of an invariable external environment by team members based on observed information. As a matter of fact, adaptation time decreases if team members observe more parameters and increases in the case of growing a priori uncertainty.

The models of team adaptation explored in this section reflect the effects of accommodation, readjustment, etc. to varying external conditions. A series of examples illustrate the processes of team adaptation to abrupt and gradual changes in external conditions.

Recall that a fundamental difference between teams and organizations is that the former possess no formal hierarchy (even despite the presence of an informal leader). For instance, consider organizations (a hierarchy makes an indispensable attribute of almost any organization, perhaps, except network organizations [22]). Here the "readjustment" problem of the principles and conditions of functioning to any variations in external conditions or other essential parameters is solved at higher hierarchical levels and the solution is subsequently "translated" to lower levels. The current section deals with the models of independent adaptation of teams to varying conditions.

Adaptation. Let us introduce a series of basic notions. Adaptation has a close correlation to self-development and self-organizing. *Self-development* is self-motion connected with transition to a higher level of organization [15]. Next, *self-motion* is a change of a certain object under the influence of its internal contradictions, factors and conditions. In this case, external actions play the modifying or mediating role.

The notion of *self-organizing* [15] (as a *process* leading to creation, reproduction or perfection of a complex system organization) seems more general. The term "a self-organizing system" was pioneered by W. Ashby [1].

Interestingly, the independent choice of functions, volumes of works, etc. by agents (see the models of teams in the previous section) can be viewed as *team's self-organizing* (in contrast to the centralized organization of activity by a Principal in hierarchical organizational systems).

Adaptation (from Latin *adaptatio* 'fit') is adjustment to the conditions of existence and habituation to such conditions; (in social systems) a type of interaction with an external environment, used to agree the requirements and expectations of its participants [15, 18]. Within the framework of the models of teams, we will comprehend adaptation as the process of changes in the actions chosen by team members based on current information in varying external conditions of team functioning (in the general case, including the functions and volumes of works).

One would easily identify several embedded levels of adaptation in any system (see Fig. 1.3), namely:

- changing of awareness about an external environment;
- changing of behavior (actions chosen using available information);

- changing of system parameters, which enables implementing more efficient behavior in varying conditions (*learning*);
- purposeful modification of an external environment (*active adaptation*).

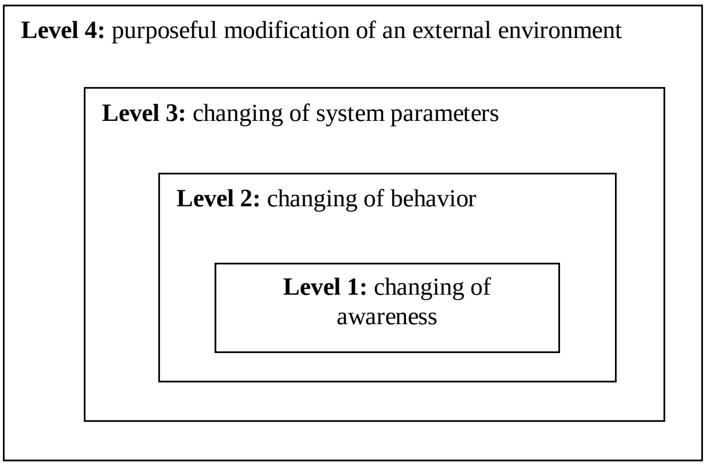

Fig. 1.3*. The levels of adaptation*

This section discusses levels 1 and 2 (level 3 of adaptation, known as *learning*, will be studied in the next section).

Control theory possesses broad experience in solving adaptive control problems for "technical" systems. Nevertheless, little knowledge has been accumulated in constructing dynamic adaptive models for socio-economic systems to date (in this context, we refer to close directions of research, *viz*., evolutionary games and Bayes models of learning, see the surveys in [4, 16, 23, etc.]).

Team members appear rational (their interests are described by goal functions, and the rational behavior of each agent presupposes maximization of its goal function). However, at each instant team members make decisions–choose their actions–under the conditions of incomplete awareness. With the course of time, they accumulate information on uncertain parameters. Different "strategies" of agents' behavior are possible depending on the goals they have.

The first variant consists in choosing at each instant such actions that facilitate rapid accumulation of maximum information on uncertain parameters (identification of their values). The *identification* stage being completed, agents then easily choose actions maximizing their goal functions. The described "behavioral strategy" agrees with the tradition of *identification theory*.

The second variant is choosing at each instant such actions that maximize the agents' payoffs within a current period, with "parallel" accumulation of information on the state of nature. We model exactly this "behavioral strategy" below.

And finally, in the third ("synthetic") variant agents choose such trajectories (sequences of actions for a given horizon) that maximize their cumulative payoff with due account of the identification effects. The corresponding models form a promising topic of future investigations.

The specifics of teams is that each agent corrects its beliefs about an uncertain parameter based on the observations of an external environment and, moreover, the observations of the actions and results of other agents (endeavoring to "explain" their choice). Fig. 1.4 demonstrates the general structure of the model of team adaptation.

Let us pass to the formal statement of the model. Consider a team $N = \{1,2, \dots, n\}$ composed of n agents. The external conditions of functioning (see the definition of adaptation above) of this team are described by the *state of nature* $\theta \in \Omega$, predetermining all essential characteristics of an external environment. Agent $i \in N$ has interval[4] information $\omega_i(\theta) \subseteq \Omega$ on the state of nature, and this information does not contradict the actual state of the things, i.e., $\forall\ \theta \in \Omega$, $\forall\ i \in N$: $\theta \in \omega_i(\theta)$.

[4] Traditionally, control theory studies models of adaptation mostly with the probabilistic uncertainties of an external environment. Applying the corresponding (and well-developed) mathematical tools to the problems of team adaptation seems to have good prospects.

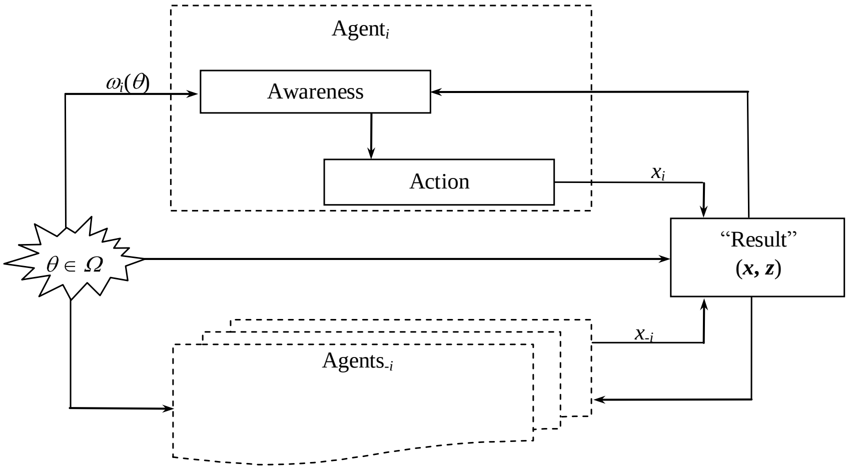

Fig. 1.4. *The model of team adaptation: the general structure*

The total result $z = G(\theta, x)$ of the team depends on the action vector $x = (x_1, x_2, \ldots, x_n) \in X' = \prod_{i \in N} X_i$ of all team members, where $x_i \in X_i$, and on the state of nature θ. We believe that each agent observes the action vector of all agents, the total result and the payoffs of all agents.

Suppose that the payoff of each agent depends on the state of nature θ and the total result z of the team: $f_i(z) = f_i(\theta, G(x, \theta))$, $i \in N$. Moreover, the set of agents N, their real-valued goal functions $\{f_i(\cdot)\}$ and feasible sets $\{X_i\}$, as well as the set Ω of the feasible states of nature, the function $G(\cdot)$ and the fact of observing the total result, all payoffs and the action vector by each team member form common knowledge.[5] If agents choose their actions simultaneously and independently, we have a game among them.

Denote by

$$E_N(\theta) = \{\{x_i\}_{i \in N} \in X' \mid \forall\, i \in N, \forall\, y_i \in X_i$$
$$f_i(\theta, G(\theta, x_1, \ldots, x_n)) \geq f_i(\theta, G(\theta, x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n))\} \quad (1.20)$$

the set of parametric Nash equilibria. Here the parameter is the state of nature–see the link between awareness and action in Fig. 1.4).

The set Ω_0 of the feasible states of nature is common knowledge among all agents. And so, by supposing that they eliminate the existing uncertainty via maximum guaranteed result evaluation, we obtain the following set of equilibria in their game:

$$E(\Omega_0) = \{\{x_i\}_{i \in N} \in X' \mid \forall\, i \in N, \forall\, y_i \in X_i$$
$$\min_{\theta \in \Omega_0} f_i(\theta, G(\theta, x_1, \ldots, x_n)) \geq \min_{\theta \in \Omega_0} f_i(\theta, G(\theta, x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n))\}. \quad (1.21)$$

Designate by $\pi(x) \subseteq \Omega$ the set of states of nature, under which the action vector observed by agents represents an equilibrium:

$$\pi(x) = \{\theta \in \Omega \mid \exists\, \Omega_0 \colon \theta \in \Omega_0, x \in E(\Omega_0)\}. \quad (1.22)$$

Let $g = (g_1, g_2, \ldots, g_n) \in \Re^n$ be the vector of the values of agents' goal functions observed by them.

Denote by

$$\delta(g, z) = \{\theta \in \Omega \mid f_j(\theta, z) = g_j, j \in N\} \quad (1.23)$$

the set of states of nature, under which (in combination with the observed result z) the observed payoffs g of agents can be implemented. We analyze the awareness of agents in a greater detail. Agent i disposes of (at most) four "sources" of information on the state of nature:

1) the a priori individual information $\omega_i(\theta) \subseteq \Omega$;

2) the actions of other agents; by observing them and assuming the rational decision-making of opponents (see the link between awareness and action in Fig. 1.4), an agent may perform reflexion (believing that

[5] Possible extensions of the model are different cases of incomplete observations; e.g., an agent does not observe the action vectors of other agents, their payoffs, etc. (see the previous section).

common knowledge takes place at the first level of the awareness structure[6] [16]), i.e., assess the information $\pi(x)$ on the state of nature leading to exactly these rational choices of the opponents;

3) the payoffs g of agents; using this information, agents make conclusions on the states of nature, under which the observed result leads to the observed payoffs–see the expression (1.23);

4) the set $\rho \subseteq \Omega$ of states of nature, under which the observed vector of agents' actions leads to the observed result z:

$$\rho(x, z) = \{\theta \in \Omega \mid G(\theta, x) = z\}. \qquad (1.24)$$

Due to the assumptions introduced earlier, information 2)–4) is common knowledge among agents. By observing same parameters, different agents should identically (and predictably for the opponents) modify their beliefs about the state of nature In other words, the following information forms common knowledge:

$$I(x, z, g) = \pi(x) \cap \rho(x, z) \cap \delta(g, z) \subseteq \Omega.$$

The last assumption, in combination with the assumption that each agent believes in common knowledge at the first level of an awareness structure, rule out agents' reflexion with respect to the awareness of their opponents. Still, this issue may be considered in further investigations.

On the basis of the above information sources, agent i can evaluate the estimate $J_i \subseteq \Omega$ of the state of nature as the intersection of the common knowledge $I(x, z, g)$ and its individual information ω_i:

$$J_i(\omega_i, x, z, g) = \omega_i \cap I(x, z, g). \qquad (1.25)$$

Let θ_0 be the actual state of nature; consider several models in the order of gradually growing complexity, namely, the single-agent model and the multi-agent model with the static and dynamic modes.[7]

The single-agent static model. Suppose that an agent makes a decision (chooses a certain action) one-time. Then, at the moment of decision-making, the agent knows only the set $\omega \subseteq \Omega$ representing the values of the states of nature. We believe that agent's decision-making in the conditions of interval uncertainty agrees with the principle of the maximum guaranteed result (MGR). This means that the agent chooses the action

$$x_{\text{MGR}}(\omega) = \arg \max_{x \in X} \min_{\theta \in \omega} f(\theta, G(\theta, x)). \qquad (1.26)$$

Recall that we focus on the static mode (one-time choice of an action by an agent) and other agents are missed. And so, the agent appears unable to use the information (1.25) on the observed result or its payoff.

Example 1.1. Set $n = 1$, $x \geq 0$, $\Omega = [1; 4]$, $\omega = [2; 4]$; $\theta_0 = 3$, $z = x / \theta$,

$$f(\theta, z) = (\theta - \alpha z)\, z - z^2 / 2, \qquad (1.27)$$

where $\alpha > 0$ is a (known) dimensional constant. Conceptually, it is possible to treat the agent as the manufacturer of some product with the demand depending on production volume. In this case, θ stands for the demand level (in terms of quantity and quality). The greater is the value of θ, the higher are the price ($\theta - \alpha\, z$) and quality requirements; to reach a same "volume," the manufacturer needs applying greater efforts (represented by action x). On the other hand, the higher is the production volume, the lower is the price.

According to the payoff function (1.27), the agent's payoff is defined as the difference between the income (the price multiplied by the production volume) and the costs described by a quadratic function.

Should the agent know the state of nature for sure, it will choose the action

$$x^*(\theta) = \frac{\theta^2}{2\alpha + 1}, \qquad (1.28)$$

maximizing the goal function which depends on the state of nature and its action:

$$f_0(\theta, x) = (\theta - \alpha\, x / \theta)\, x / \theta - x^2 / (2\, \theta^2). \qquad (1.29)$$

As far as the goal function (1.27) increases monotonically in θ for any feasible actions of the agent, the expression (1.26) implies that

$$x_{\text{MGR}}(\omega) = 4 / (2\, \alpha + 1). \qquad (1.30)$$

By observing (10) and the result $x_{\text{MGR}}(\omega) / \theta_0$ or the payoff $f(\theta_0, x_{\text{MGR}}(\omega) / \theta_0)$ (or even both these quantities), the agent can uniquely evaluate the actual state of nature θ_0. •

[6] More complicated cases are also possible, where agents have nontrivial mutual awareness. Then the parametric Nash equilibrium (1.20) is replaced by the informational equilibrium in the game of agents.

[7] In the discrete model studied, the "static" mode corresponds to one-time choice of actions by agents, whereas the "dynamic" mode answers to a sequence of such choices.

Example 1.1 illustrates situations when one-time observation of corresponding information is sufficient for retrieving the actual state of nature by the agent. Interestingly, repeating the observations and employing auxiliary information on the choice of other agents (if available) becomes useless. However, sometimes one-time observation appears insufficient for the agent. We provide an example.

Example 1.2. Set $n = 1$, $x \geq 0$, $z = x$, $\theta = (\theta_p, \theta_c)$ $\Omega = [1; 4] \times [1; 4]$, $\omega = [2; 4] \times [2; 4]$; $\theta_0 = (3; 3)$,

$$f(\theta, x) = (\theta_p - \alpha x)\, x - x^2\, \theta_c / 2\,, \qquad (1.31)$$

where $\alpha \geq 0$ is a (known) dimensional constant. In contrast to Example 1.1, the state of nature represents a two-dimensional vector (the first component describes price parameters, while the second one stands for costs' parameters).

Should the agent know the state of nature for sure, it will choose the action

$$x^*(\theta) = \frac{\theta_p}{2\alpha + \theta_c}. \qquad (1.32)$$

As far as the goal function (1.31) increases monotonically in θ_p and decreases monotonically in θ_c for any feasible actions of the agent, the expression (1.26) dictates that

$$x_{\text{MGR}}(\omega) = 1 / (\alpha + 2). \qquad (1.33)$$

In Example 1.2, the agent's action coincides with its result; hence, observation of its actual payoff is the only information source for the agent. Based on this observation, it draws the following conclusion regarding the set of feasible states of nature:

$$I = \{\theta \in \Omega \mid \theta_c = 2\ \theta_p\ (\alpha + 2) - 6\ \alpha - 9\}. \qquad (1.34)$$

For instance, under $\alpha = 1$ it appears from (1.25) that

$$J = \{(\theta_p; \theta_c) \mid \theta_c = 6\ \theta_p - 15,\ \theta_p \in [17/6; 19/6]\}. \qquad (1.35)$$

We emphasize that still the agent's information and the actual situation are consistent, i.e., $J \subseteq \omega$ and $\theta_0 \in J$, $\theta_0 \in I$. •

The single-agent dynamic model. The "repeated" usage of information obtained during observation of actions becomes possible when the agent chooses actions many times. Assume that the agents choose their actions simultaneously at each step and the steps are "uniform."

Example 1.3. In the conditions of Example 1.2, set $\alpha = 1$ and imagine that the agent makes decisions sequentially several times. After the first "step," it possesses the information (1.35). According to the expression (1.26), the agent will choose the action $x_{\text{MGR}}(J) = 17 / 31$ at the second "step." Observing its payoff as the result of this action, the agent may uniquely retrieve the actual state of nature $\theta_0 = (3; 3)$.

Thereby, in the present example, two observations (two "steps") are sufficient for retrieving the missed information by the agent. •

We omit the case of multiple agents in the static mode, proceeding directly to the dynamic mode.

The multi-agent dynamic model. Denote by $x_i^t \in X_i$ the action of agent i at the moment t and by $x^{1,\,t}$ the set of action vectors of all agents during t periods. By the end of the period t, the information

$$I(x^t, z^t, g^t) = \pi(x^t) \cap \rho(x^t, z^t) \cap \delta(g^t, z^t) \subseteq \Omega$$

is the common knowledge of all agents.

Based on all information sources, agent i can evaluate the estimate $J_i^t \subseteq \Omega$ of the state of nature within t periods as the intersection of the common knowledge $I(x^t, z^t, g^t)$ and its individual information J_i^{t-1} corresponding to the previous period:

$$J_i^t = J_i^{t-1} \cap I(x^t, z^t, g^t). \qquad (1.36)$$

In other words, its estimate of the state of nature is narrowed to the set

$$J_i^t(\omega_i, x^{1,\,t}, z^{1,\,t}, g^{1,\,t}) = \omega_i \cap \bigcap_{\tau=1}^{t} I(x^\tau, z^\tau, g^\tau). \qquad (1.37)$$

Example 1.4. Consider the Cournot's oligopoly model [13, 14] in uncertain conditions.

Set $n = 2$, $x_i \geq 0$, $i = 1, 2$, $z = x_1 + x_2$, $\Omega = [1; 5]$, $\omega_1 = [1; 4]$; $\omega_2 = [2; 5]$; $\theta_0 = 3$,

$$f_i(\theta, z) = (\theta - \alpha z)\, z - x_i^2\, r / 2, \qquad (1.38)$$

where $\alpha > 0$, $r > 0$ are known dimensional constants. That is, agents differ only in their awareness about the state of nature.

Should agents know the state of nature for sure, they will choose the actions

$$x_i^*(\theta) = \frac{\theta}{4\alpha + r}, \; i = 1, 2. \qquad (1.39)$$

Since the payoff functions (1.38) increase monotonically in θ for any feasible actions of agents, in the first period the agents choose the actions

$$x_1^1 = 1 / (4\alpha + r), \; x_2^1 = 2 / (4\alpha + r) \qquad (1.40)$$

according to the expression (1.26).

As the result of such choice, agents retrieve the actual state of nature by observing the action vectors and payoffs several times. •

Let us introduce the notion of "*adaptation time*" in a team. This is the time required for the unambiguous identification of the state of an invariable external environment by team members based on observed information. Adaptation time (the duration of a corresponding transient process) depends on the parameters observed by agents, the dimensionality of the vector describing the states of nature and the properties of point-set mappings (1.22)-(1.24). Adaptation time equals 1 (one period) in Examples 1.1 and 1.4 and 2 (two periods) in Example 1.3.

As a matter of fact, adaptation time decreases (more specifically, does not increase) if team members observe more parameters and increases (more specifically, does not decrease) in the case of higher dimensionality of the vector of states of nature and higher a priori uncertainty (wider sets $\{\omega_i\}$ describing the individual information of agents).

Example 1.5. Consider Example 1.4 and add agent 3 with the initial awareness $\omega_3 = [2.5; 3.5]$.

If each agent still observes the actions and payoffs of all agents, then all agents can retrieve the state of nature during one step, exactly as in Example 1.4. Adaptation time may increase if the awareness of agents "gets worse", i.e., the set of observed parameters is narrowed or only a few aggregate attributes remain observable (e.g., the sum of actions performed by all agents).

Therefore, suppose that the agent i observes its action x_i and payoff g_i, as well as the total action z of all agents[8] z. Moreover, the fact of such observations makes the common knowledge of all agents.

Under known x_i, z and g_i, the equation

$$(\theta - \alpha z)\, z - x_i^2\, r / 2 = g_i$$

possesses a unique solution θ, $i = 1, 2$. In other words, increasing the number of agents would not raise adaptation time in the current example. •

Example 1.6. Keeping in mind the conditions of Example 1.5, assume that each agent observes its own action and payoff only. Then observations lead to the following equation for agent i:

$$(\theta - \alpha (x_1 + x_2))\,(x_1 + x_2) - x_i^2\, r / 2 = g_i, \qquad (1.41)$$

with two unknown variables x_{3-i} and θ, $i = 1, 2$.

Suppose that each agent considers the situation of common knowledge, i.e., the opponents are provided with the same level of awareness as the agent in question. Then an agent thinks that an opponent would choose a same action as it does. Recall that in the current example agents differ in their awareness about the state of nature. Substituting the actual payoff of the agent and

$$x_{3-i} = x_{\text{MGR}i}(\omega_i)$$

into (1.41), we obtain

$$(\theta - 2\,\alpha\, x_i^1)\, 2\, x_i^1 - (x_i^1)^2\, r / 2 = (\theta_0 - \alpha\,(x_1^1 + x_2^1))\,(x_1^1 + x_2^1) - (x_i^1)^2\, r / 2. \qquad (1.42)$$

Thus, agent i can evaluate the lower estimate of the state of nature by the end of the first period:

$$\theta_i^1 = (\theta_0 - \alpha\,(x_1^1 + x_2^1))\,(x_1^1 + x_2^1) / 2\, x_i^1 + 2\,\alpha\, x_i^1. \qquad (1.43)$$

Here $i = 1, 2$.

Take the case of $\alpha = r = 1$. Subsequently,

$$x_1^1 = 0.2, \; x_2^1 = 0.4, \; \theta_1^1 = 4, \; \theta_2^1 = 2.6.$$

---

[8] In the case of two agents, each of them can evaluate the action of another based on the known total action and its own action. For three and more agents, such information becomes insufficient for retrieving unambiguously the actions of the opponents.
но.

During the second period, agents substitute the corresponding estimates θ_1^1 and θ_2^1 into the expression (1.39). In other words, they choose the actions

$$x_1^2 = 0.8,\ x_2^2 = 0.52,$$

substitute them into the analog of formula (1.42), calculate the new estimates of the state of nature, and so on.

Generally, the estimates of the state of nature obtained by agents have the following dynamics (compare this result with (1.42)):

$$\theta_i^t = (\theta_0 - \alpha\,(x_1^t + x_2^t))\,(x_1^t + x_2^t)\,/\,2\ x_i^t + 2\,\alpha\, x_i^t,\ i = 1, 2,\ t = 1, 2, \dots . \qquad (1.44)$$

Based on these estimates, agents choose the actions (see the expression (1.39))

$$x_i^t(\theta_i^{t-1}) = \frac{\theta_i^{t-1}}{4\alpha + r},\ i = 1, 2,\ t = 1, 2, \dots . \qquad (1.45)$$

Consequently, team adaptation is described by the system of iterated functions (1.44)-(1.45), with initial conditions (1.40) defined under the a priori information of agents according to the principle of maximum guaranteed result.

Figs. 1.5 and 1.6 demonstrate the dynamics of the estimated states of nature obtained by agents (the first level of adaptation) and the dynamics of agents' actions (the second level of adaptation).

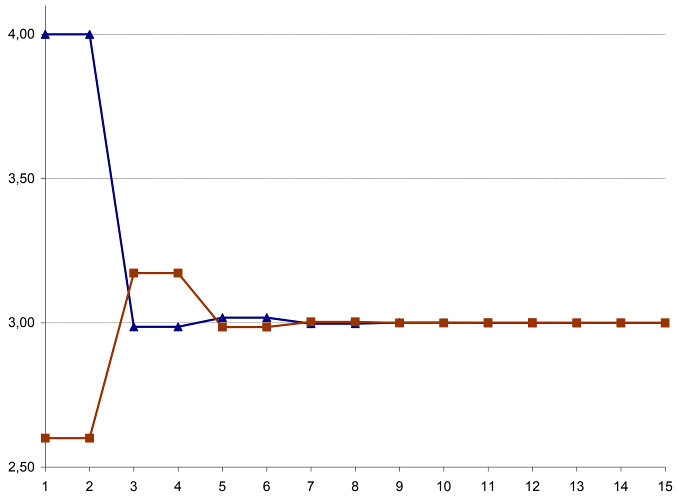

Fig. 1.5. *The dynamics of the estimated states of nature*
(*triangles–agent 1, square boxes–agent 2*)

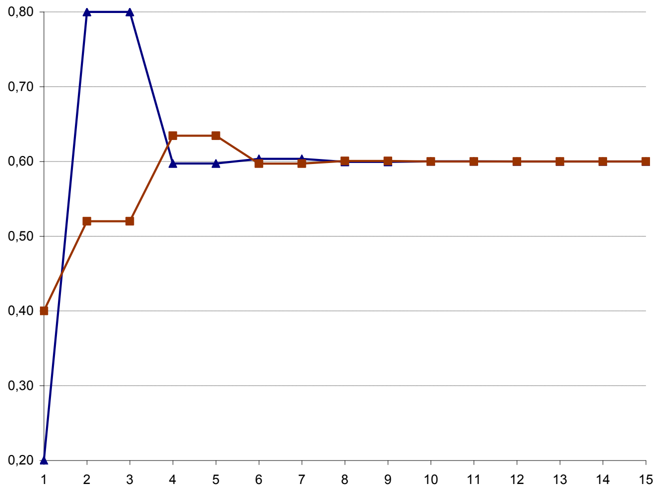

Fig. 1.6. *The dynamics of agents' actions*
(*triangles–agent 1, square boxes–agent 2*)

Obviously, the processes describing variations in the estimates obtained by agents have rapid convergence (after 8–10 steps, the variations become negligible). Interestingly, these processes tend to the actual state of nature. Furthermore, even having different a priori awareness, agents eventually choose identical actions (it seems natural, since agents have same goal functions). Strictly speaking, here adaptation time is infinite, though the period of reaching any preassigned nonempty neighborhood of the actual state of nature appears finite. •

Adaptation corresponds to accommodation, getting accustomed to varying external conditions, etc. In this section, we have discussed adaptation models allowing description of such effects (see examples in [19]). Adaptation makes sense if adaptation time does not exceed the characteristic time of external environment variation. On the other hand, external conditions may vary gradually and a team should adapt to such "slow" changes in the conditions of functioning.

Finally, note the following aspect. By the above assumption, each agent assigns same awareness to the opponents as it has. This assumption can be violated by considering more complicated structures of agents' awareness [16] and believing that agents choose informational equilibrium actions. In addition, more complicated structures of agents' "observations" may take place. For instance, some agents observe only few parameters (e.g., the actions and payoffs for a certain set of agents), while the others observe other parameters (e.g., the actions and payoffs for another set of agents and information on the state of nature), etc. Perhaps, all such cases can be described similarly to the ones discussed above.

In this section, adaptation has been treated as accommodation to certain (basically, external) conditions and as getting accustomed to them; in fact, adaptation depends on information on these conditions[9] being available to agents at the moment of their decision-making. In contrast, modification of team parameters (the third level of adaptation in Fig. 3) can be viewed as *learning.*[10] Therefore, our analysis switches to the models of team learning.

1.3. Team learning

In this section, we consider the models of iterative learning implemented during professional activity, i.e., the problem of optimal collective learning (task allocation among team members [17, 20].

While acting jointly, the members of a team or collective (*agents*) consciously or unconsciously gain experience in individual and collective activity. In other words, the process of their *learning* takes place. Here and in the sequel, we understand learning as "the process and result of gaining individual experience" [10, 20]. This interpretation is a special case of the more general notion of learning as the process of acquiring knowledge, skill, and habits. Let us consider a series of models describing the learning effects of team members during their collective work. Starting from the general problem statement and quantitative description of learning process, we consider the model of individual learning process, and then the model of learning process for a collective of agents.

General problem statement and the model of learning process. In qualitative terms, the general problem of optimal learning can be formulated as follows. Each agent in a collective is characterized by some initial level of skill (e.g., labour productivity). With the course of working activity, the labour productivity of an agent grows owing to gaining experience, improving practical skills, etc. (*learning during working*, *learning-by-doing* takes place). However, the rate of this growth (the so-called *learning rate* formally defined below) is individual for each agent. We are interested in optimal allocation of work tasks among agents. Imagine an agent with a low level of initial professional skills; being strongly loaded by work from the very beginning, this agent can improve its skills rapidly and work with higher productivity later. On the other hand, is it reasonable to load agents with higher initial professional skills? The answers to these questions are not trivial. Moreover, we have to determine what is understood as the "optimal" allocation of work

[9] Of course, generally adaptation of a system presumes not only changing the awareness and behavior of its elements (the first and the second levels of adaptation), but also changing system parameters (the third level of adaptation), e.g., the types of agents, in response to varying external conditions. Moreover, active adaptation can be considered when the system directly influences on an external environment (the fourth level of adaptation).

[10] Learning and adaptation are close phenomena. However, learning may take place in invariable external conditions, whereas adaptation happens only when these conditions vary.

tasks among agents. Possible efficiency criteria are the total costs of agents, the time required for performing a given volume of work by a collective, the result achieved within a fixed period, and so on.

Let us proceed to formal analysis. We begin with the simplest model and gradually increase the complexity. Our discussion gets confined to the case of *iterative learning* [20] corresponding to routine activities. Iterative learning is multiple reiterations of actions, trials, attempts, etc., by a system for achieving a fixed goal in invariable external conditions. Iterative learning (IL) underlies formation of human habits, conditioned reflexes of animals, learning of many technical (materialized) and cybernetic (abstract logical) systems. This is the subject of research in pedagogical and engineering psychology, psychophysiology, pedagogics, control theory, and other sciences (see the survey [20]).

The invariability of external conditions and the goal allows describing IL quantitatively via *learning curves* representing the level of learning as functions of time and the number of iterations.

According to numerous experimental data, the most important general regularity of iterative learning consists in slowed-down asymptotic behaviour of learning curves. They are monotonic; the rate of change in learning level decreases in time; the curve itself asymptotically tends to some limit. In most cases, iterative learning curves can be approximated by exponential curves.

The following two aspects of learning seem relevant. The first aspect concerns the results of learning. While learning, a system has to achieve a required result with the quality of actions and admissible costs of time, energy, etc. The second aspect is associated with the process of learning and includes adaptation of a system learnt to some activity by working (e.g., exercises), etc. Accordingly, there exist the efficiency characteristics of iterative learning and adaptation [20]. As a rule, adaptation characteristics relate to the physiological components of activity (fatiguability, etc.). This section considers just the characteristics of learning efficiency (adaptation characteristics have quite different dynamics).

Recall that iterative learning is often described by slowed-down asymptotic learning curves admitting approximation by the *exponential curves*

$$r(t) = r^{\infty} + (r^0 - r^{\infty})\, e^{-\gamma t},\ t \geq 0, \qquad (1.46)$$

or the discrete sequence[11]

$$r^k = r^{\infty} + (r^0 - r^{\infty})\, e^{-\gamma k},\ k = 1, 2, \ldots \qquad (1.47)$$

Here t is the time moment of learning, k indicates the number of iterations (trials, attempts) from the starting moment of learning; $r(t)$ (r_k) means *agent's type* (the level of practical or professional skills) at the moment t (or iteration k); $r^0 > 0$ specifies *the initial professional skills* (the type value at the starting moment of learning, i.e., the initial moment of time); r^{∞} gives the "final" value, $r^{\infty} \geq r^0$; γ is a nonnegative constant defining the rate of change in the type (*the learning rate*).

Learning of a single agent. First, we consider *the model of learning* for a single agent. Denote by $y^k \geq 0$ *the volume of work* performed by the agent within period k. If the agent's type (skill level) $r^k \in [0; 1]$ is interpreted as the share of its successful actions, then the agent achieves *the result* $z^k = r^k\, y^k$ by performing the volume of work y^k within period k.

Consequently, the agent's result (the total volume of works successfully performed by it within k periods) makes up

$$Z^k = \sum_{l=1}^{k} r^l y^l\ . \qquad (1.48)$$

On the other hand, the agent performs a greater volume of (successful and unsuccessful) works:

$$Y^k = \sum_{l=1}^{k} y^l\ . \qquad (1.49)$$

This volume of works can be treated as the "*experience*" gained by the agent [21], i.e., the "effective time" of the agent (the period consumed for learning process). By substituting (1.48) into the exponent (1.46), we have

$$r^k = 1 - (1 - r^0)\exp(-\gamma\, Y^{k-1}),\ k = 2, 3, \ldots$$

Let $y^{1,\tau} = (y^1, y^2, \ldots, y^{\tau})$, $\tau = 1, 2, \ldots$ and assume that $y^0 = 0$. Combining (1.46)-(1.49) yields the following expressions for the volumes of works performed successfully and the types of the agent:

[11] Throughout this section, the superscript denotes the number of a time period, whereas the subscript designates the number of an agent. In the single-agent model, the subscript is omitted.

$$Z^k = \sum_{l=1}^{k} y^l \{1-(1-r^0)\exp(-\gamma \sum_{m=1}^{l-1} y^m)\}, \qquad (1.50)$$

$$r^k = 1 - (1 - r^0) \exp(-\gamma \sum_{l=1}^{k-1} y^l), k = 2, 3, \dots \qquad (1.51)$$

Under a fixed total volume of works, the agent's type is uniquely defined by formula (1.51) regardless of the distribution of works over time periods. Therefore, the problem of maximizing the agent's type with a fixed total volume of works makes no sense within the framework of this model.

The model incorporates three "macro-parameters," namely, the total volume of works Y, the number of periods T, and the result Z. The desired variable is the "*learning trajectory*" $y^{1,T}$.

Optimal learning problems lie in extremalization of one variable under fixed other variables.[12] Thus and so, the following problem statements appear reasonable:

1. Fix the total volume of works Y to-be-performed by the agent and the result Z to-be-achieved. It is necessary to find a trajectory minimizing the time of achieving the result:

$$\begin{cases} T \to \min \\ Y^T \le Y \\ Z^T \ge Z \end{cases}. \qquad (1.52)$$

The problem (1.52) can be called the minimum time problem.

2. Fix the total volume of works Y to-be-performed by the agent and the learning time T. It is necessary to find a trajectory maximizing the result Z:

$$\begin{cases} Z^\tau \to \max \\ Y^\tau \le Y \\ \tau \le T \end{cases}. \qquad (1.53)$$

The problem (1.53) can be called *the problem of optimal agent learning*. Interestingly, this problem is closest to pedagogical problems: under fixed time and the volume of teaching material, one has to distribute this material in time to maximize the "volume of learnt material" (the "quality of learning"). However, didactic aspects (the content of teaching material) seem insignificant due to the routine character of learning.

Since the expression (1.50) is monotonic in the sum of the volume of agent's works and the learning period, the problem (1.53) can be rewritten as

$$\sum_{l=1}^{T} y^l \exp(-\gamma \sum_{m=1}^{l-1} y^m) \to \min_{\{y^{1,T} | \sum_{\tau=1}^{T} y^\tau = Y\}}. \qquad (1.54)$$

Formula (1.54) does not employ the initial skill r^0 of the agent. In other words, we have the following result.

Assertion 1 [17]. The solution to the optimal learning problem is independent from the initial skill of the agent.

This fact is interesting for the methodology of learning. Really, only the differences between the individual learning rates of independent agents are essential from the viewpoint of the ultimate results.

3. Fix the learning time T and the result Z to-be-achieved by the agent. It is necessary to find a learning trajectory minimizing the total volume of works:

$$\begin{cases} Y^\tau \to \min \\ \tau \le T \\ Z^T \ge Z \end{cases}. \qquad (1.55)$$

Any of the problems (1.52)-(1.55) admits reduction to appropriate dynamic programming problems.

Learning of multiple agents. Let us generalize the derived results to the case of several agents working simultaneously. First, we study the situation when the results and type of each agent are independent from the results and types of other agents. And second, we switch to the problem of learning of dependent agents.

[12] In more general case, one would desire to extremize some functional (e.g., learning expenses, learning quality, etc.) taking into account some additional constraints, varying several variables simultaneously, etc. All these problems form the prospective subject of future research.

Consider *a team* which is a set $N = \{1, 2, \dots, n\}$ of n agents. By analogy to the expressions (1.50) and (1.51), we obtain the following formulas for the volumes of works performed successfully and the types of agents, respectively:

$$\sum_{l=1}^{k} y_i^l \{1-(1-r_i^0)\exp(-\gamma_i \sum_{m=1}^{l-1} y_i^m)\}, \qquad (1.56)$$

$$r_i^k = 1 - (1 - r_i^0) \exp(-\gamma_i \sum_{l=1}^{k-1} y_i^l), \; k = 2, 3, \dots, \; i \in N. \qquad (1.57)$$

Suppose that the result of the team makes the sum of the individual results of all team member:

$$Z^k = \sum_{i=1}^{n} Z_i^k, \; k = 1, 2, \dots \qquad (1.58)$$

In this case, *the problem of optimal learning of a collective* (compare with (1.53)) is defined by

$$Z^T \to \max_{\{y_i^{1,T} | \sum_{\tau=1}^{T}\sum_{i=1}^{N} y_i^\tau = Y\}}, \qquad (1.59)$$

which is equivalent to

$$\sum_{i=1}^{n}\sum_{l=1}^{T} y_i^l \{1-(1-r_i^0)\exp(-\gamma_i \sum_{m=1}^{l-1} y_i^m)\} \to \max_{\{y_i^{1,T} | \sum_{\tau=1}^{T}\sum_{i=1}^{N} y_i^\tau = Y\}}. \qquad (1.60)$$

The problem (1.60) can be solved by the dynamic programming method. The optimal solution to the problem (1.60) generally depends on the individual rates of the agents' learning $\{\gamma_i\}$ and their initial skills $\{r_i^0\}$.

Assertion 2 [17]. Let the learning rates of agents be identical. Then the optimal distribution of works consists in performing the whole volume of works by an agent with the maximum initial skill. If the initial skills of the agent coincide, then the optimal distribution of works is performing the whole volume of works by an agent with the maximum learning rate.

Thus, in the case when all agents have same learning rates, the solution to the optimal learning problem turns out "degenerate"; only one agent works and learns while the rest do nothing. On the other hand, such collectives would be hardly considered to full value. Unfortunately, such situations happen in real life.

Now, suppose that agents differ both in their initial skills and learning rates.

Nominally, the solution structure of the problem (1.60) when the whole volume of works is performed by the "best" agent (in the sense of the initial skill and learning rate) includes more variables under a single constraint. Substantially, the problem may have other constraints besides the one imposed on the total volume of works performed by team members. The most natural constraint applies to the maximum volume of works performed by each agent within one iteration (period of time).

Collective learning. Up to this point, our discussion of agents' learning-by-doing has been based on the following assumption. Each agent learns only owing to "its own experience." Nevertheless, collectives are remarkable for experience exchange; agents gain additional experience by observing the activity of other agents (their successes and difficulties). To take this effect into account, we describe the "experience" gained by an agent as the sum of its own actions and the weighted sum of actions of other agents. Subsequently, we have the following expressions for the volumes of works performed successfully and agents' types, respectively:

$$Z_i^k = \sum_{l=1}^{k} y_i^l \{1-(1-r_i^0)\exp(-\gamma_i \sum_{j=1}^{n}\alpha_{ij}\sum_{m=1}^{l-1} y_j^m)\}, \qquad (1.61)$$

$$r_i^k = 1 - (1 - r_i^0) \exp(-\gamma_i \sum_{j=1}^{n}\alpha_{ij}\sum_{l=1}^{k-1} y_j^l), \; k = 2, 3, \dots, \; i \in N. \qquad (1.62)$$

Here the constants $\{\alpha_{ij} \geq 0\}$ can be interpreted as *the efficiencies of experience transfer* from agent j to agent i ($i, j \in N$).

Then the optimal learning problem takes the form

$$\sum_{i=1}^{n}\sum_{l=1}^{T} y_i^l \{1-(1-r_i^0)\exp(-\gamma_i \sum_{j=1}^{n}\alpha_{ij}\sum_{m=1}^{l-1} y_i^m)\} \to \max_{\{y_i^{1,T} | \sum_{\tau=1}^{T}\sum_{i=1}^{N} y_i^\tau = Y\}}. \qquad (1.63)$$

Example 1.7. Consider the problem (1.63) in the case of two agents with $T = 11$, $r_1^0 = 0.1$, $r_2^0 = 0.3$, $\gamma_1 = \gamma_2 = 0.75$, $Y = 10$ and $||\alpha_{ij}|| = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Clearly, both agents have identical learning rates, but agent 2 has higher initial skill. Agent 1 learns by its own experience and the experience of agent 2 (even more efficiently than by its own experience). And agent 2 learns only by its own experience. The dynamics of the agents' types is demonstrated by Fig. 1.7. The dynamics of the optimal volumes of works can be found in Fig. 1.8.

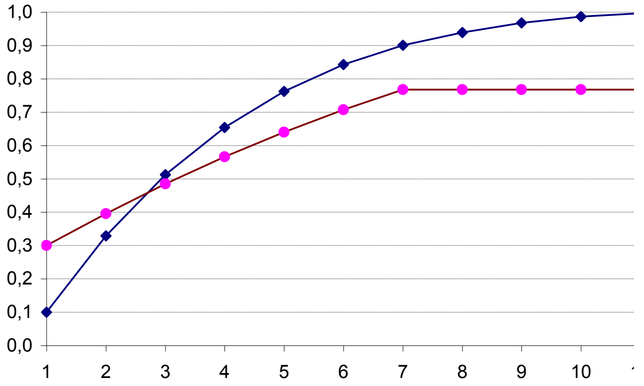

Fig. 1.7. *The dynamics of agents' types*

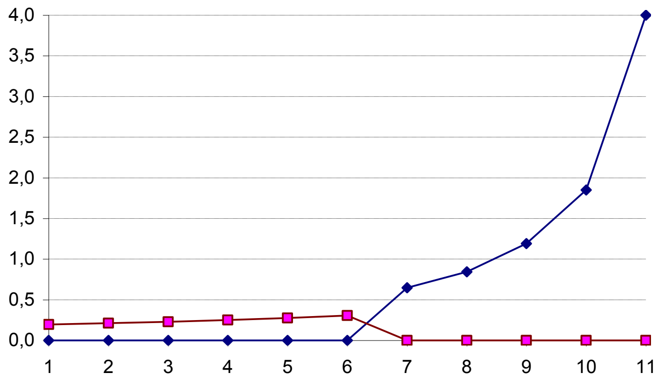

Fig. 1.8. *The dynamics of the optimal volumes of works*

Within the starting six periods, agent 1 does not work but "observes" the actions of agent 2. However, the professional skill of agent 1 grows more rapidly than that of agent 2. Beginning with period 7, the optimal solution is performing the whole volume of works by agent 1 instead of agent 2.

This example elucidates how the lack of initial skill can be compensated owing to the efficient learning by other's experience. In another interpretation, agent 2 acts as a teacher, a tutor, an instructor possessing a higher initial skill and training agent 1. At some moment, the learner "outruns" the teacher and works independently.

Discussion. We have considered the models of learning-by-doing. Under the assumption that the volume of works already performed by an agent reflects the "experience" gained by it, we have stated and solved the optimal learning problem of allocating the volumes of works performed by agents in certain time intervals. The conducted analysis brings to the following conclusions:

- with a fixed total volume of works of one agent, the characteristics of the learning efficiency are independent from the volumes of works distributed over time periods;
- the solution to the problem of optimal iterative learning of one agent appears independent from its initial skill;
- the higher is agent's learning rate, the greater volume of works should be performed by it within last periods (and, respectively, the smaller volume of works should be assigned to the starting periods to improve its initial skill);
- the optimal learning strategy consists in increasing the volume of agent's works with the course of time; the higher is the learning rate, the more "convex" is the optimal learning trajectory;
- in the absence of constraints on the individual volumes of works, the whole volume of works in a team should be done by the "best" agent (in the sense of initial skill and learning rate);
- the lack of agent's initial skill can be compensated via efficient learning using its own experience and the experience of other agents.

In conclusion, we acknowledge the existence of learning curves with more complex structures (than exponential or logistic ones), the so-called sequential logistic curves corresponding to the development of various adjacent or more complex kinds of activity, generalized logistic curves, etc. Their detailed analysis goes beyond the scope of this paper. Although, if the learning laws of team members are known (even despite their complexity), then the problem of optimal distribution of works can be stated similarly. Yet, searching for the general (preferably, analytical) solution to this problem makes the subject of future research.

REFERENCES

1. Ashby W. An Introduction to Cybernetics. – London: Chapman and Hall, 1956. – 295 p.
2. Aumann R. Agreeing to Disagree // The Annals of Statistics. 1976. Vol. 4. No. 6. P. 1236–1239.
3. Burkov V., Goubko M., Korgin N., Novikov D. Introduction to Theory of Control in Organizations. – London: CRC Press, 2015. – 342 p.
4. Camerer C. Behavioral Game Theory: Experiments in Strategic Interactions. – Princeton: Princeton University Press, 2003. – 544 p.
5. Chkhartishvili A., Novikov D. Information Equilibrium: Punctual Structures of Information Distribution // Automation and Remote Control. 2003. Vol. 64. No. 10. P. 1609–1619.
6. Chkhartishvili A., Novikov D. Models of Reflexive Decision-Making // Systems Science. 2004. Vol. 30. No. 2. P. 45–59.
7. Chkhartishvili A., Novikov D. Stability of Information Equilibrium in Reflexive Games // Automation and Remote Control. 2005. Vol. 66. No. 3. P. 441–448.
8. Halverson C., Tirvizi S. Effective Multicultural Team: Theory and Practice. – Heidelberg: Springer, 2013. – 292 p.
9. Holmstrom B. Moral Hazard in Teams // Bell Journal of Economics. 1982. Vol. 13. P. 324–340.
10. Hull C. Principles of Behavior and Introduction to Behavior Theory. – Oxford: Appleton Century Company, 1943. – 422 p.
11. Kim K., Roush F. Team Theory. – Chichester: Ellis Horwood, 1997. – 246 p.
12. Marchak J., Radner R. Economic Theory of Teams. – New Haven – London: Yale University Press, 1976. – 345 p.
13. Mas-Collel A., Whinston M. D., Green J. R. Microeconomic Theory. – N.Y.: Oxford University Press, 1995. – 981 p.
14. Myerson R. Game Theory: Analysis of Conflict. – London: Harvard University Press, 1991. – 568 p.
15. Novikov A., Novikov D. Research Methodology: from Philosophy of Science to Research Design. – Amsterdam: CRC Press, 2013. – 200 p.
16. Novikov D., Chkhartishvili A. Reflexion and Control: Mathematical Models. – London: CRC Press, 2014. – 298 p.
17. Novikov D. Collective Learning-by-doing / Proceedings of IFAC Symposium on Advances in Control Education. – Nizhny Novgorod, 2012. P. 408–412.
18. Novikov D. Control Methodology. – N.Y.: Nova Science Publishing, 2013. – 76 p.
19. Novikov D. Mathematical Models of Teams Building and Functioning. – Moscow: Fizmatlit, 2008. – 184 p. (in Russian)
20. Novikov D. Regularities of Iterative Learning. – Moscow: ICS RAS, 1998. – 101 p. (in Russian)
21. Novikov D. Team Building under Pareto Uncertainty. Proceedings of the 17th IFAC World Congress. – Seoul, 2008. – P. 1633–1638.
22. Novikov D. Theory of Control in Organizations. – N.Y.: Nova Science Publishing, 2013. – 341 p.
23. Weibull J. Evolutionary Game Theory. – Cambridge: MIT Press, 1996. – 265 p.