

# Multi-time-scale Markov decision processes for organizational decision-making

Christian Wernz

Received: 12 September 2012 / Accepted: 13 October 2013 / Published online: 8 November 2013  
© Springer-Verlag Berlin Heidelberg and EURO - The Association of European Operational Research Societies 2013

**Abstract** Decision-makers in organizations and other hierarchical systems interact within and across multiple organizational levels and take interdependent actions over time. The challenge is to identify incentive mechanisms that align agents' interests and to provide these agents with guidance for their decision processes. To this end, we developed a multiscale decision-making model that combines game theory with multi-time-scale Markov decision processes to model agents' multi-level, multi-period interactions. For a two-level, multi-period, multi-agent problem, we derived closed-form solutions that determine optimal incentive levels, provide agents with decision rules, and identify information needs. We show the difference between a single-time-scale model, where agents make decisions with the same frequency, and a multi-time-scale model, where superior agents make decisions less frequently than their subordinates. Our results show that in the single-time-scale case, agents can make optimal decisions with information merely from the current and subsequent period. In the multi-time-scale case, data from the entire decision horizon are necessary. This paper contributes to multiscale decision theory by generalizing solutions for multi-period interactions that were previously limited to a few time periods. We apply our model to a management problem with an account manager and multiple customer representatives.

**Keywords** Multiscale decision theory · Markov decision processes · Hierarchical games · Stochastic games · Management application

**Mathematics Subject Classification** 90C40 · 90B50 · 91A65 · 91A15 · 91A35

---

C. Wernz (✉)  
Grado Department of Industrial and Systems Engineering, Virginia Tech,  
Blacksburg, VA 24061, USA  
e-mail: cwernz@vt.edu

## 1 Introduction

Decision-making over multiple periods is a challenging task for agents. The decision process is further complicated when agents across multiple hierarchical levels affect each other's decisions and outcomes, and when uncertainties need to be considered. We model and analyze the interactions between decision-making agents across two hierarchical levels, where the higher-level agent relies on cooperative decisions by multiple lower-level agents. A conflict of interest exists between the two levels as the preferred actions of the lower-level agents reduce the superior agent's likelihood to achieve its goal. To align interests, the higher-level agent can offer a share of its reward as an incentive to the lower-level agents. A sufficiently large incentive results in cooperative behavior by the lower-level agents. In this paper, we determine agents' optimal decisions and incentives given their organizational interdependencies and multi-period interactions.

Decisions by higher and lower-level agents occur on different time scales. The higher-level agent makes decisions on a slower time scale, i.e., less frequently, than lower-level agents. In other words, the decisions of the higher-level agents are more strategic, whereas the decisions by the lower-level agents are more operational.

To analyze this type of interaction and to determine optimal decisions for agents, we combine a game-theoretic model with multi-time-scale Markov decision processes (MMDPs) to account for the multi-period, multi-time-scale interdependencies.

This paper contributes to multiscale decision theory (MSDT) by generalizing solutions for multi-period interactions that were previously limited to few periods. In general, analytical models have been limited to either multi-agent *one-period* problems, or *single-agent* multi-period problems. Multi-agent, multi-period problems have not yet been solved analytically. This paper develops a unified model for simultaneously solving hierarchical and temporal decision challenges. Lastly, this paper makes a contribution to the literature on MMDPs (e.g., Chang et al. 2003) by proposing a novel modeling approach and formulation technique.

The paper is organized as follows: Sect. 2 discusses the related literature and Sect. 3 introduces the model. In Sect. 4, the agent interaction is analyzed, first for a single-time-scale situation and then for the multi-time-scale case. Section 5 provides a numerical example of the model. Section 6 presents the conclusions.

## 2 Literature review

We review the literature on hierarchical and organizational decision-making that has influenced the development of MSDT. We discuss the state-of-the-art methods in MSDT and show their relationship with the literature on multi-time-scale modeling and multi-organizational scale modeling.

### 2.1 Hierarchical and organizational decision-making

The phenomenon of organizations and their hierarchical structures was first studied by sociologists and economists. In his 1922 posthumously published work on

“Economy and Society”, Max Weber (Weber 1978) views bureaucracies as hierarchical structures composed of rational decision-makers. The seminal work by Williamson (1979) on transaction-cost economics linked the existence of organizations and their hierarchical structures to transaction costs associated with asset specificity, uncertainty, and transaction frequency. Transaction costs are the reasons why organizations are formed as opposed to single consumer–producer interactions in the market place. Williamson (1967, 1970) had earlier developed a quantitative model to determine the optimal number of hierarchical levels in organizations.

Hierarchy within organizations can be explained with information-processing limitations and social preferences of humans. Radner (1992) views decentralization as a necessity to overcome the limitations of individuals in processing information, but attributes the fact that decentralization takes the form of hierarchy to sociological and psychological reasons rather than pure economic reasons. The concept of bounded rationality reconciled economists’ rationality assumption with behavioral observations of the limitations of human decision-making. Bounded rationality is most prominently associated with the work of Simon (1957, 1972), who received the Noble Memorial Prize in Economics for his work in this area. Earlier discussions of the bounded rationality concept can be traced back to “Foundations of Statistics” Savage (1954).

Rational decision-making of interdependent individuals in organization can be formulated mathematically using cooperative and non-cooperative game theory. The theory of teams by Marschak (1955) and Marschak and Radner (1972) formulates the problems of collaborating individuals who have the same interests and beliefs, but do not share the same information. In contrast to this cooperative game theory approach, non-cooperative game theory is used to describe conflicts of interests between individuals in organizations. Personal and/or team objectives can be in conflict with that of other individuals, teams, or the higher-level objectives, particularly in large organizations. Incentives can be used to align conflicting interests and motivate work that supports higher-level, and thus more important, organizational objectives (Wernz 2008).

This paper determines effective incentives for objective alignment, and our model has similarities to principal-agent models. Principal-agent theory describes a hierarchical interaction between two decision-makers with incomplete and asymmetric information. The superior decision-maker (principal) offers a contract to the subordinate decision-maker (agent). The contract is designed to motivate the agent to perform work in the interest of the principal (Milgrom and Roberts 1992). Although a number of papers, e.g., Besanko (1985), Plambeck and Zenios (2000), discuss principal-agent models for multi-periods, also called dynamic principal agent models, no publication has yet discussed the multi-time-scale aspect in which the principal makes decisions on a strategic level, i.e. less frequently than the subordinate agent.

Organizational agent interactions have also been studied in fields outside of economics, particularly in operations research, management science, computer science, and systems engineering. In the following paragraphs, we present contributions that have most significantly influenced the development of MSDT.

A taxonomy to classify and formally describe hierarchical agent interactions was developed by Schneeweiss (1995, 2003a, b). A unified notation for various hierarchical agent interaction situations was proposed. The author acknowledges the different time scales present in organizational systems (strategic, tactical, operational), but the models only describe one-period interactions.

Mesarovic et al. (1970) developed a mathematical framework to describe and analyze multi-level hierarchical systems, but again only for single period interactions. A mathematical theory of coordination was developed based on the conceptual and formal aspects of hierarchical systems. From their work, we adopted the interaction relationship between lower and higher levels, in which the success of the supramal agent depends on the performance of the infimal agent.

To account for uncertainty and temporal dynamics in strategic agent interactions, stochastic game theory was developed (Shapley 1953). Stochastic games are a combination of Markov decision processes (MDPs) (Puterman 1994) and classical game theory. Shapley (1953) was the first to propose an algorithm that solves stochastic games. Zachrisson (1964) coined the term “Markov games” to emphasize the connection to MDPs. For survey papers on stochastic games and solution algorithms see Howard (1960), Pollatschek and Avi-Itzhak (1969), Raghavan and Filar (1991), Mertens (1992), Filar and Vrieze (1996), and Bowling and Veloso (2000). We are using an MDP formulation in our model to account for the stochastic, multi-period agent interactions. In the following section, we further discuss this topic with an emphasis on multi-time-scale aspects.

## 2.2 Multi-time-scale decision-making models

Researchers have used various approaches to model multi-time-scale phenomena via MDPs. One can distinguish between two classes of contributions: (1) multi-time-scale systems that are modeled with MDPs and (2) one-time-scale systems with imposed hierarchical, multi-time-scale structures. The latter is mostly used to increase the computational efficiency of solution algorithms; e.g., Sutton (1995), Hauskrecht et al. (1998), Parr (1998), and Sutton et al. (1999). Our model falls into class (1) of multi-time-scale systems, and the related literature is discussed in the following paragraphs.

Chang et al. (2003) proposed MMDPs. The authors’ model is an extension of Muppala et al. (1996) and Goseva-Popstojanova and Trivedi (2000), in which the lower level is modeled as a Markov chain and the upper level is described via a Markov reward process. MMDP has been applied to production planning in semiconductor fabrication (Panigrahi and Bhatnagar 2004; Bhatnagar and Panigrahi 2006), management of hydro-power plants (Zhu et al. 2006), and has influenced work on target tracking in sensor networks (Yeow et al. 2005, 2007) and reverse supply chains (Wongthatsanekorn et al. 2010). Chang (2004) extended his earlier work to include game-theoretic interactions between agents with conflicting interests. In Chang’s model, agents make their decisions sequentially. In our model, we assume that agents do not know the actions taken by the other agents, which we recognize by applying a non-cooperative game-theoretic model with simultaneous decision-making.

Jacobson et al. (2003) proposed periodically time-inhomogeneous Markov decision processes (PTMDPs). The evolution of the system is described by  $(N + 1)$ -periodic sequences of reward functions and transition probabilities. The first  $N$  epochs are fast-scale epochs, while the interval  $N + 1$  is a slow-scale cycle. Their approach presents an alternative to our approach of modeling multi-time-scale evaluation, yet without including strategic agent interactions.

In their book, Sethi and Zhang (1994) focused on hierarchical decision-making in stochastic manufacturing systems. Different hierarchical levels have different time scales associated with their decisions. Hierarchy is discussed from a temporal perspective, but organizational interactions across hierarchies are not explicitly modeled.

Delebecque and Quadrat (1981) analyzed a multi-time scale phenomenon in hydroelectric power generation with singularly perturbed MDPs. In singularly perturbed MDPs, small perturbations on fast time scales affect the aggregate behavior observable on slower time scales. Singularly perturbed MDPs have been extensively studied in the literature; see Yin and Zhang (1998, 2005) for an overview. In our model, fast scale decisions also affect higher-level phenomena, but the link is through organizational hierarchies, not perturbations.

### 2.3 Multiscale decision theory

The model that describes the type of hierarchical agent interaction considered in this paper was introduced by Wernz and Deshmukh (2007a, b). Rewards are passed down from a supramal agent to motivate cooperative decisions by infimal agents. A cooperative decision by an infimal agent improves the supramal agent's chances of success, i.e., increases the probability of a transition to a state associated with higher rewards.

Wernz and Deshmukh (2010) built upon this initial model in their development of the multiscale decision-making framework. Their model was the first comprehensive approach that could derive closed-form, analytic solutions for one period, multi-organizational scale interactions in large, hierarchical organizations. The multiscale decision-making framework has been applied to a two-level management challenge (Wernz and Deshmukh 2007b), a two-level production planning problem (Wernz and Deshmukh 2007a), a three-level service operations application (Wernz and Henry 2009), and a three-stage supply chain problem (Henry and Wernz 2013). Further details on these applications are presented at the end of the example problem in Sect. 3.1.

Wernz and Deshmukh (2012) extended a three-level model from one to three periods. The next step in the theory's development is to extend the multiscale decision-making model to account for interactions over many periods. The first efforts in this direction were documented in conference proceedings by Wernz and Deshmukh (2009, 2010). This paper extends their work by generalizing the model from two agents to  $n + 1$  agents, by formulating theorems and developing proofs, by applying the model to a challenge problem, by providing a numerical example, and by discussing the results and their implications in a decision processes context.

### 3 Model

We motivate our model with an example problem from the customer service industry. The purpose of the example is to illustrate the applicability of the multiscale decision-making model and to put the mathematical result into a decision-relevant context.

#### 3.1 Example problem

In a customer service division of an organization, an account manager is supported by a number of customer representatives. The manager is responsible for renewing customer contracts at regular time intervals. The day-to-day customer care is handled by the representatives, who answer customer phone calls to solve customer problems.

When contacting customers for service contract renewals, the manager decides between two approaches: (1) she calls the customer to discuss new contract terms, or (2) sends a small gift along with a letter with details of the new terms. Both actions have the same costs for the organization. As a result of the manager's action, customers either renew or discontinue their contracts. Both outcomes are possible for a given action, but a call is more likely to result in a renewal than when a letter with a gift is sent.

The representatives, when receiving a customer's phone call with a service request, can also take one of two actions. To address a customer request, representatives can (1) be strict and insist on the contractual agreement when identifying resources to solve the customer's problem, or (2) be lenient and use additional resources to try to satisfy the customer. As a result of the interaction, the customer is either satisfied or dissatisfied. Satisfying the customer is associated with additional cost to the organization. A lenient response to a customer request most likely results in a satisfied, high-cost customer, while a strict response most likely results in an unsatisfied, low-cost customer. Still, both outcomes are possible for a given action.

The representative's pay depends on the cost he produces; lower costs are preferred and result in a higher salary from the organization. Consequently, the representative prefers to be strict. However, since customer satisfaction has an influence on contract renewal rates, the manager might be willing to compensate the payoff difference for the representatives. If it is financially advantageous for her, the manager will offer an incentive to the representatives. We assume that the incentive is a proportional share of the manager's pay. The higher the customer renewal rate, the higher the manager's pay, and the higher the incentive for the representatives.

A number of questions arise for the manager and the representatives: how high does the incentive for the representatives have to be to switch from the initially preferred action of being strict to the cooperative action of being lenient? What is the maximum share of her pay the manager is willing to offer before the cost of incentives exceeds the benefits? What role do the transition probabilities that describe the link between actions and outcomes play? How can the manager and the representatives compute their optimal decisions? What information do they need to

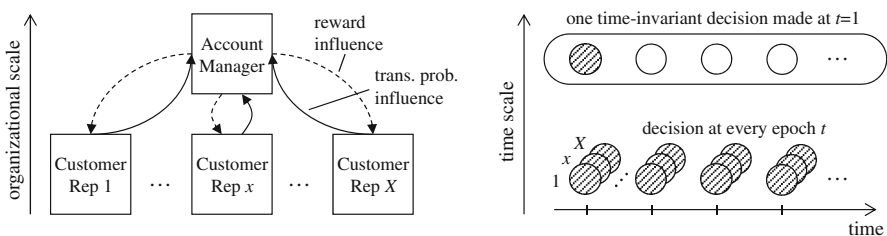
determine an optimal decision? How do the temporal dynamics of representatives interacting multiple times with a customer before contract renewal affect all of these aspects?

The temporal dynamics will be modeled via MDPs. To account for the fact that the manager only makes decisions from time to time, while the representatives make their decisions more frequently, we model the manager’s decision as time-invariant until the next contract renewal phase. This means that the manager’s decision (i.e., to call or send a letter) and the incentive level she offers to the representatives do not change over time. This difference in decision frequency accounts for the multi-time-scale nature of the problem, and the fact that the time-scale is coupled to the organizational scale.

Figure 1 illustrates the model and the relationship between organizational scale and time scale. On the left side of Fig. 1, a dependency graph (Dolgov and Durfee 2004) depicts the mutual influence of the decision-makers. Corresponding to the organizational scale, the right side of Fig. 1 shows the multi-period and multi-time-scale aspects across both hierarchical levels.

The following section will describe the example problem mathematically and in general terms. The mathematical model description is applicable to a wide range of hierarchical decision-making problems. The customer service example serves the purpose of illustrating results and providing the context for the decision process studied.

As mentioned in the literature review, similar examples with applications in production planning, maintenance service operations, and supply chain management have been analyzed. In production planning, a hierarchical chain consisting of operations manager, production planner, and material handler have conflicts of interest related to profitability, meeting production deadlines, and inventory costs (Wernz and Deshmukh 2007a, 2010b). In a three-level maintenance service scenario, an account manager uses incentives to align the interests of her subordinate maintenance supervisor, who in turn seeks to align the actions of his maintenance workers (Wernz and Henry 2009; Wernz and Deshmukh 2012). Multiscale decision-making has also been used to coordinate the decisions between organizations. In a supply chain, a retailer, a wholesaler, and a producer can use incentives to coordinate their chain-wide decisions (Henry and Wernz 2013). Further applications in health care and hospital management are currently being explored (Sudhaakar and Wernz 2013).



**Fig. 1** Multiscale model for two hierarchical levels and many periods

### 3.2 Model formulation

In the general model, the account manager will be referred to as the supremal agent, or agent SUP, and the representatives as infimal agents or agents INF1, INF2 etc., generally INF $x$ , with  $x = 1, \dots, X$ .

We formulate the multi-period model as a discrete time problem with  $N$  epochs that demarcate  $N - 1$  periods following the MDP notation of Puterman (1994). Time begins with decision epoch 1, the start of period 1, and ends with period  $N - 1$  followed by the final epoch  $N$ .

At every decision epoch  $t = 1, \dots, N - 1$ , agents INF $x$  carry out actions  $a_{n(x),t}^{INFx}$ , where  $n(x)$  is the index that denotes the chosen action from agent INF $x$ 's action space. No action is carried out at the final epoch  $N$ . Agent SUP decides in the first decision epoch to take an action indexed by  $m$  for this and all future periods, i.e.,  $a_{m,t}^{SUP}$  with  $t = 1, \dots, N - 1$ . By limiting agent SUP to only one decision that governs the entire time horizon, we account for the different decision frequencies between agent SUP and agents INF $x$ . In our example, the manager makes one decision (call customer or send letter), while the representatives have multiple interactions and make a new decision (strict or lenient) each time.

Associated with each agent are states  $s_{i,t}^{SUP} \in \mathcal{S}_t^{SUP}$  and  $s_{k(x),t}^{INFx} \in \mathcal{S}_t^{INFx}$  for every period. Index  $i$  and index function  $k(x)$  denote the agents' possible states given period  $t$ . Depending on their current states  $s_{i,t}^{SUP}$  and  $s_{k(x),t}^{INFx}$ , and their actions  $a_{m,t}^{SUP}$  and  $a_{n(x),t}^{INFx}$ , agents move to states  $s_{j,t+1}^{SUP}$  and  $s_{l(x),t+1}^{INFx}$  with probability  $p_t^{SUP}(s_{j,t+1}^{SUP} | s_{i,t}^{SUP}, a_{m,t}^{SUP})$  and  $p_t^{INFx}(s_{l(x),t+1}^{INFx} | s_{j(x),t}^{INFx}, a_{n(x),t}^{INFx})$  in the following period. The state to which the agent transitions determines the agent's reward (referred to as "pay" in the example), which is notated by  $r_t^{SUP}(s_{j,t+1}^{SUP})$  and  $r_t^{INFx}(s_{l(x),t+1}^{INFx})$ , respectively. This process repeats for every period  $t = 1, \dots, N - 1$ .

We analyze a situation in which each agent has a distinct set of two actions and two states in each period. The action spaces for agents SUP and INF $x$  at a given decision epoch  $t$  are denoted by  $\mathcal{A}_t^{SUP} := \{a_{1,t}^{SUP}, a_{2,t}^{SUP}\}$ ,  $\mathcal{A}_t^{INFx} := \{a_{1,t}^{INFx}, a_{2,t}^{INFx}\}$  with  $t = 1, \dots, N - 1$  and their state spaces by  $\mathcal{S}_t^{SUP} := \{s_{1,t}^{SUP}, s_{2,t}^{SUP}\}$ ,  $\mathcal{S}_t^{INFx} := \{s_{1,t}^{INFx}, s_{2,t}^{INFx}\}$  with  $t = 1, \dots, N$ .

In the context of our example, action  $a_{1,t}^{SUP}$  corresponds to the manager's phone call, and  $a_{2,t}^{SUP}$  to sending a letter to the customer. The corresponding outcomes are  $s_{1,t}^{SUP}$  for contract renewal and  $s_{2,t}^{SUP}$  for no contract renewal. For the customer representatives,  $a_{1,t}^{INFx}$  corresponds to being lenient and  $a_{2,t}^{INFx}$  to being strict. The outcomes are  $s_{1,t}^{INFx}$ , a satisfied, high-cost customer, and  $s_{2,t}^{INFx}$  a dissatisfied, low-cost customer.

The initial rewards for agents SUP and INF $x$  are

$$r_t^{SUP}(s_{1,t+1}^{SUP}) := \rho_{1,t}^{SUP}, \quad r_t^{SUP}(s_{2,t+1}^{SUP}) := \rho_{2,t}^{SUP}, \tag{1}$$



$$r_t^{\text{INF}x}(s_{1,t+1}^{\text{INF}x}) := \rho_{1,t}^{\text{INF}x}, \quad r_t^{\text{INF}x}(s_{2,t+1}^{\text{INF}x}) := \rho_{2,t}^{\text{INF}x} \tag{2}$$

The data can alternatively be represented in vector notation:

$$R_t^{\text{SUP}} := \begin{pmatrix} \rho_{1,t}^{\text{SUP}} \\ \rho_{2,t}^{\text{SUP}} \end{pmatrix}, \quad R_t^{\text{INF}x} := \begin{pmatrix} \rho_{1,t}^{\text{INF}x} \\ \rho_{2,t}^{\text{INF}x} \end{pmatrix} \tag{3}$$

The initial state-dependent transition probabilities for agent SUP are

$$p_t^{\text{SUP}}(s_{1,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{1,t}^{\text{SUP}}) := \alpha_{i,1,t}^{\text{SUP}}, \quad p_t^{\text{SUP}}(s_{2,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{1,t}^{\text{SUP}}) := 1 - \alpha_{i,1,t}^{\text{SUP}} \tag{4}$$

$$p_t^{\text{SUP}}(s_{1,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{2,t}^{\text{SUP}}) := 1 - \alpha_{i,2,t}^{\text{SUP}}, \quad p_t^{\text{SUP}}(s_{2,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{2,t}^{\text{SUP}}) := \alpha_{i,2,t}^{\text{SUP}} \tag{5}$$

with  $i = 1, 2, m = 1, 2$  and  $0 \leq \alpha_{i,m,t}^{\text{SUP}} \leq 1$ . In matrix notation, we can compactly represent the data of agent SUP as

$$P_t^{\text{SUP}}(s_{i,t}^{\text{SUP}}) := \begin{pmatrix} \alpha_{i,1,t}^{\text{SUP}} & 1 - \alpha_{i,1,t}^{\text{SUP}} \\ 1 - \alpha_{i,2,t}^{\text{SUP}} & \alpha_{i,2,t}^{\text{SUP}} \end{pmatrix}. \tag{6}$$

The rows of the matrices correspond to the actions and the columns correspond to the states to which the agent transitions. For agents INFx, the notation can be represented accordingly, and in matrix notation the transition probabilities are

$$P_t^{\text{INF}x}(s_{j(x),t}^{\text{INF}x}) := \begin{pmatrix} \alpha_{j(x),1,t}^{\text{INF}x} & 1 - \alpha_{j(x),1,t}^{\text{INF}x} \\ 1 - \alpha_{j(x),2,t}^{\text{INF}x} & \alpha_{j(x),2,t}^{\text{INF}x} \end{pmatrix} \tag{7}$$

with index function  $j(x) = 1, 2$ .

The hierarchical interactions between agent SUP and agents INFx consist of bottom-up and a top-down influences. The bottom-up influence is the effect agents INFx have on agent SUP's expected outcome. The states to which agents INFx transition influence agent SUP's chances of success, i.e., its transition probability  $p_t^{\text{SUP}}$ . MSDT models this influence using an additive influence function  $f_{x,t}$ . The final transition probability of agent SUP is based on the initial transition probability and the influence functions such that

$$p_{final,t}^{\text{SUP}}(s_{k,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{m,t}^{\text{SUP}}, (s_{l(x),t+1}^{\text{INF}x})_{1,\dots,X}) = p_t^{\text{SUP}}(s_{k,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{m,t}^{\text{SUP}}) + \sum_{x=1}^X f_{x,t}(s_{k,t+1}^{\text{SUP}} | s_{l(x),t+1}^{\text{INF}x}) \tag{8}$$

for indices  $i, k, m, l(x) = 1, 2$ , and with  $(s_{l(x),t+1}^{\text{INF}x})_{1,\dots,X} := s_{l(1),t+1}^{\text{INF}1}, s_{l(2),t+1}^{\text{INF}2}, \dots, s_{l(X),t+1}^{\text{INF}X}$ . We choose the influence function to be a constant and define it as

$$f_{x,t}(s_{k,t+1}^{\text{SUP}} | s_{l(x),t+1}^{\text{INF}x}) := \begin{cases} c_{x,t} & \text{if } k = l(x), \\ -c_{x,t} & \text{if } k \neq l(x), \end{cases} \quad \text{with } c_{x,t} > 0. \tag{9}$$

Constant  $c_{x,t}$  is referred to as the *change coefficient*. We denote the aggregate

influence of all agents  $INF_x$  in period  $t$  as  $C_t$  with  $C_t := \sum_{x=1}^X f_{x,t} \left( s_{k,t+1}^{SUP} | s_{l(x),t+1}^{INF_x} \right)$ . Since probabilities can neither be negative nor exceed unity,  $0 \leq p_{final,t}^{SUP}(\cdot) \leq 1$  must hold, which bounds the aggregate influence coefficient to

$$0 \leq C_t \leq \min \left\{ \alpha_{i,1,t}^{SUP}, \alpha_{i,2,t}^{SUP}, 1 - \alpha_{i,1,t}^{SUP}, 1 - \alpha_{i,2,t}^{SUP} \right\}. \tag{10}$$

The meaning of the chosen change coefficient structure in (9) is as follows: in each period, states  $s_{1,t}^{INF_x}$  increase the probability of state  $s_{1,t}^{SUP}$ , and reduce the probability of state  $s_{2,t}^{SUP}$  accordingly. The effect on agent SUP’s transition probabilities change in opposite direction for agents  $INF_x$ ’s states  $s_{2,t}^{INF_x}$ , i.e., state  $s_{2,t}^{SUP}$  becomes more likely and state  $s_{1,t}^{SUP}$  becomes less likely. This effect on transition probabilities applies to situations where the states of infimal agents lead to agent SUP reaching a specific state with higher or lower probability. The aggregate change coefficient  $C_t$  describes the influence of all infimal agents combined. In the context of our example, representatives who can achieve high customer satisfaction contribute to a higher contract renewal rate for the manager.

The top-down influence of the hierarchical interaction between agent SUP and agents  $INF_x$  is an incentive payment by agent SUP to all agents  $INF_x$ . Each agent  $INF_x$  receives a share  $b_{x,t}$  of agent SUP’s reward. We refer to  $b_{x,t}$  as the *share coefficient*. The final reward in period  $t$  for an agent  $INF_x$  is its initial reward plus the incentive, i.e.,

$$\begin{aligned} r_{final,t}^{INF_x} \left( s_{k,t+1}^{SUP}, s_{l(x),t+1}^{INF_x} \right) &:= r_t^{INF_x} \left( s_{l(x),t+1}^{INF_x} \right) + b_{x,t} \times r_t^{SUP} \left( s_{k,t+1}^{SUP} \right) \\ &= \rho_{l(x),t}^{INF_x} + b_{x,t} \times \rho_{k,t}^{SUP}. \end{aligned} \tag{11}$$

Agent SUP’s initial reward is reduced by the reward share given to agents  $INF_x$ . Agent SUP’s final reward is

$$r_{final,t}^{SUP} \left( s_{k,t+1}^{SUP} \right) := \left( 1 - \sum_{x=1}^X b_{x,t} \right) \times \rho_{k,t}^{SUP}. \tag{12}$$

Figure 2 provides a graphical summary of the model. It shows the interaction between agent SUP and an agent  $INF_x$  over two periods. The following assumptions about the model’s parameters are made:

$$\rho_{1,t}^{SUP} > \rho_{2,t}^{SUP}, \quad \rho_{1,t}^{INF_x} < \rho_{2,t}^{INF_x} \tag{13}$$

$$\alpha_{i,m,t}^{SUP}, \alpha_{j(x),n,t}^{INF_x} > \frac{1}{2} \tag{14}$$

$$\alpha_{1,1,t}^{SUP} > \alpha_{2,1,t}^{SUP}, \quad \alpha_{1,1,t}^{INF_x} > \alpha_{2,1,t}^{INF_x}. \tag{15}$$

Inequalities in (13) express that agent SUP prefers state  $s_{1,t+1}^{SUP}$  over  $s_{2,t+1}^{SUP}$  and agents  $INF_x$  initially prefer  $s_{2,t+1}^{INF_x}$  over  $s_{1,t+1}^{INF_x}$ . This corresponds to the manager’s preference for contract renewal and the representatives’ goal of low costs. One can see that a conflict of interest exists between agent SUP and agents  $INF_x$ . Agents  $INF_x$  initially

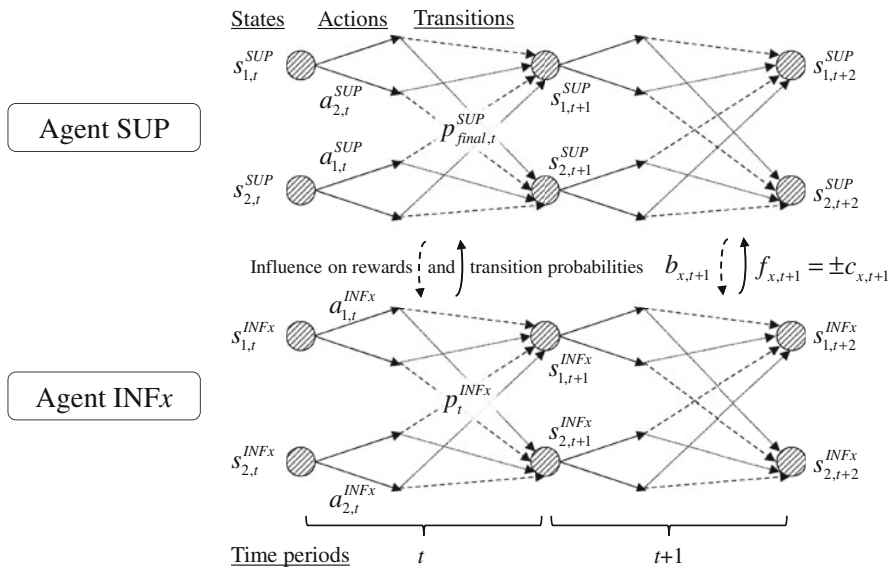


Fig. 2 Schematic model representation

prefer to reach state  $s_{2,t+1}^{INFx}$ , which would reduce agent SUP’s chance of attaining its preferred state  $s_{1,t+1}^{SUP}$ .

Expression (14) states that an action is linked to the state with the same index. In other words, there is an index-corresponding action for every state, which is the most likely consequence of the respective action. This restriction circumvents redundant cases in the analysis.

Inequalities in (15) express that, given action  $a_{1,t}$ , the transition probability  $\alpha_{1,1,t}$  from state  $s_{1,t}$  to  $s_{1,t+1}$  is greater than  $\alpha_{2,1,t}$ , which denotes the probability of transitioning from state  $s_{2,t}$  to  $s_{1,t+1}$ . That means it is more likely to remain in state 1 than to switch from state 2 to state 1, given the corresponding action  $a_{1,t}$ . Applied to our example, this implies that the probability of contract renewal is greater than non-renewal when the customer has planned to renew and the manager makes the phone call. At the representative level, this means that a satisfied customer will more likely remain satisfied as opposed to dissatisfied when the representative is lenient.

Agents seek to maximize their rewards. For each agent INFx in period  $t$  the expected reward is:

$$\begin{aligned}
 & E \left( r_{final,t}^{INFx} | s_{i,t}^{SUP}, (s_{j(x),t}^{INFx})_{1,\dots,X}, a_{m,t}^{SUP}, (a_{n(x),t}^{INFx})_{1,\dots,X} \right) := \\
 & \sum_{k=1}^2 \sum_{l(1)=1}^2 \dots \sum_{l(x)=1}^2 \dots \sum_{l(X)=1}^2 \left[ r_{final,t}^{INFx} (s_{k,t+1}^{SUP}, s_{l(x),t+1}^{INFx}) \times P_{final,t}^{SUP} (s_{k,t+1}^{SUP} | s_{i,t}^{SUP}, a_{m,t}^{SUP}, (s_{l(x),t+1}^{INFx})_{1,\dots,X}) \right. \\
 & \left. \times \prod_{x=1,\dots,X} P_t^{INFx} (s_{l(x),t+1}^{INFx} | s_{j(x),t}^{INFx}, a_{n(x),t}^{INFx}) \right].
 \end{aligned}
 \tag{16}$$

The expected reward for agent SUP can be calculated in the same way by replacing

$$r_{\text{final},t}^{\text{INFx}} \left( s_{k,t+1}^{\text{SUP}}, s_{l(x),t+1}^{\text{INFx}} \right) \text{ with } r_{\text{final},t}^{\text{SUP}} \left( s_{k,t+1}^{\text{SUP}}, \left( s_{l(x),t+1}^{\text{INFx}} \right)_{1,\dots,X} \right) \text{ in (16).}$$

Agents seek to maximize the sum of all  $r_{\text{final},t}$  over the time horizon, i.e., their cumulative rewards. The cumulative rewards from period  $t$  to period  $N - 1$  for agents INFx and SUP are

$$r_{\text{final}(t)}^{\text{INFx}} \left( s_{i,t}^{\text{SUP}}, s_{j(x),t}^{\text{INFx}} \right) := \sum_{\tau=t}^{N-1} r_{\text{final},\tau}^{\text{INFx}} \left( s_{k,\tau+1}^{\text{SUP}}, s_{l(x),\tau+1}^{\text{INFx}} \right) \tag{17}$$

$$r_{\text{final}(t)}^{\text{SUP}} \left( s_{i,t}^{\text{SUP}}, \left( s_{j(x),t}^{\text{INFx}} \right)_{1,\dots,X} \right) := \sum_{\tau=t}^{N-1} r_{\text{final},\tau}^{\text{SUP}} \left( s_{k,\tau+1}^{\text{SUP}}, \left( s_{l(x),\tau+1}^{\text{INFx}} \right)_{1,\dots,X} \right) \tag{18}$$

To calculate the *expected* cumulative reward, which agents need to calculate to determine their optimal course of action, the backward induction principle (Bellman 1957) is applied, starting in the last period of the time horizon and working backwards to period 1. The expected cumulative reward from period  $t$  to period  $N - 1$  is

$$\begin{aligned} & E \left( r_{\text{final}(t)}^{\text{INFx}} | s_{i,t}^{\text{SUP}}, \left( s_{j(x),t}^{\text{INFx}} \right)_{1,\dots,X}, a_{m,t}^{\text{SUP}}, \left( a_{n(x),t}^{\text{INFx}} \right)_{1,\dots,X} \right) \\ & := E \left( r_{\text{final},t}^{\text{INFx}} | s_{i,t}^{\text{SUP}}, \left( s_{j(x),t}^{\text{INFx}} \right)_{1,\dots,X}, a_{m,t}^{\text{SUP}}, \left( a_{n(x),t}^{\text{INFx}} \right)_{1,\dots,X} \right) \\ & \quad + \sum_{k=1}^2 \sum_{l(1)=1}^2 \dots \sum_{l(x)=1}^2 \dots \sum_{l(X)=1}^2 P_{\text{final},t}^{\text{SUP}} \left( s_{k,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, a_{m,t}^{\text{SUP}}, \left( s_{l(x),t+1}^{\text{INFx}} \right)_{1,\dots,X} \right) \\ & \quad \times \prod_{x=1,\dots,X} p_t^{\text{INFx}} \left( s_{l(x),t+1}^{\text{INFx}} | s_{j(x),t}^{\text{INFx}}, a_{n(x),t}^{\text{INFx}} \right) \times E^* \left( r_{\text{final}(t+1)}^{\text{INFx}} | s_{k,t+1}^{\text{SUP}}, \left( s_{l(x),t+1}^{\text{INFx}} \right)_{1,\dots,X} \right) \end{aligned} \tag{19}$$

for  $i, j, m, n = 1, 2, t = 1, \dots, N - 1$  and with  $E^* \left( r_{\text{final}(N)}^{\text{INFx}} | s_{k,N}^{\text{SUP}}, \left( s_{l(x),N}^{\text{INFx}} \right)_{1,\dots,X} \right) = 0$ .

In the following analysis, we determine optimal incentive levels and optimal actions for the agents.

### 4 Analysis

We assume that agents are risk-neutral and rational, i.e., agents maximize their expected utilities, or equivalently, their expected rewards. Rational agents are able to calculate both their own expected rewards, as well as the other agents' expected rewards. They can decide which decisions yield the highest expected rewards for themselves. Hence, agents will engage in a game-theoretic reasoning process, recognizing the dependency of each other's decisions. The Nash equilibrium concept is used to determine which decisions the rational agents take for a given incentive.

The game occurs in two stages. In the first stage, agent SUP determines the incentive level, i.e., the share coefficient. In the second stage, agents take the respective actions in response to the incentive level chosen. To determine an optimal action, agents evaluate their and the other agents' cumulative expected rewards of the entire time horizon, i.e.,  $E\left(r_{\text{final}(1)}^{\text{SUP}}|\cdot\right)$  and  $E\left(r_{\text{final}(1)}^{\text{INF}x}|\cdot\right)$ .

The Nash equilibrium is the set of agents' actions that are the best responses to other agents' preferred actions. Expressed mathematically, this means that the Nash equilibrium is a strategy profile  $a^* = \left(a_{m,t}^{\text{SUP}*}, a_{n(1),t}^{\text{INF}1*}, \dots, a_{n(X),t}^{\text{INF}X*}\right)_{t=1,\dots,N-1}$  for a given  $b^*$  such that

$$E\left(r_{\text{final}(1)}^{\text{SUP}}|\cdot, a^*\right) \geq E\left(r_{\text{final}(1)}^{\text{SUP}}|\cdot, a^-\right) \text{ and} \tag{20}$$

$$E\left(r_{\text{final}(1)}^{\text{INF}x}|\cdot, a^*\right) \geq E\left(r_{\text{final}(1)}^{\text{INF}x}|\cdot, a^-\right) \forall x \tag{21}$$

for all  $a^-$ , where  $a^-$  refers to any strategy profile different from  $a^*$ .

It is easy to show that agent SUP has a dominant action  $a_{1,t}^{\text{SUP}}$ , i.e., a best response regardless of other agents' actions. Since agent SUP's preferred state is  $s_{1,t}^{\text{SUP}}$ , it will choose action  $a_{1,t}^{\text{SUP}}$  as it has the highest probability of reaching state  $s_{1,t}^{\text{SUP}}$ . This result holds for all levels of incentives. For a formal proof, see Wernz and Henry (2009).

Agent SUP's chances of success, i.e., reaching states  $s_{1,t}^{\text{SUP}}$ , are improved when agents INF $x$  switch from their initially preferred actions  $a_{2,t}^{\text{INF}x}$  to  $a_{1,t}^{\text{INF}x}$ . For a sufficiently high level of incentives, agents INF $x$  will choose the cooperative actions  $a_{1,t}^{\text{INF}x}$ . Agent SUP will offer this incentive if its expected gain in reward is larger than the cost of the incentive.

In the following sections, we will determine the conditions for which a collaborative Nash equilibrium exists. First, we determine the Nash equilibrium for a single-time-scale model, where agents make decisions at the same frequency, before analyzing the multi-time-scale case. The single-time-scale model provides important results that are the basis for multi-time-scale analysis. In the multi-time-scale model, hierarchical interactions occur on multiple time scales, i.e., agent SUP and agents INF $x$  make decisions at different frequencies.

### 4.1 Single time-scale model

In the single-time-scale model, agent SUP can vary its decision from period to period and can select different incentive levels (share coefficients) in each period. In contrast, agent SUP can choose only one decision and one share coefficient for all periods in the multi-time-scale case.

We begin the analysis by determining the share coefficients  $b_{x,t}$  for which agents INF $x$  would switch from non-cooperative to cooperative actions. Using backward induction, we start in the final decision epoch, and then move to earlier periods.

**Theorem 1** *The share coefficients  $b_{x,N-1}$  of the final period  $N - 1$  that are cost-minimal for agent SUP and that motivate agents INFx to choose the cooperative actions  $a_{1,N-1}^{INFx}$  are*

$$b_{x,N-1}^* = \frac{\rho_{2,N-1}^{INFx} - \rho_{1,N-1}^{INFx}}{2c_{x,N-1} \left( \rho_{1,N-1}^{SUP} - \rho_{2,N-1}^{SUP} \right)}. \tag{22}$$

*Proof* see Appendix.

The result of Theorem 1 is surprising insofar as although all agents INFx affect the expected reward of agent SUP, only data from the respective infimal agent play a role in the optimal share coefficient. Furthermore, neither the agents' prior states nor the transition probabilities affect the result. Transition probabilities and prior states do not affect the share coefficient, because at the optimal incentive level  $b_{x,N-1}^*$  agents INFx are indifferent between their actions (still, weakly prefer the cooperative action) and the transition probability variables cancel each other out.

Next, we will investigate whether agent SUP is willing to give the share  $b_{x,N-1}^*$  of its reward to agents INFx in return for their cooperation. The so-called *participation condition* checks if agent SUP receives a higher expected reward if an incentive according to  $b_{x,N-1}^*$  is offered to agents INFx. Mathematically, the participation condition is

$$\begin{aligned} & E \left( r_{\text{final}(t)}^{SUP} \middle| (b_{x,t})_{1,\dots,x} s_{i,t}^{SUP}, (s_{j(x),t}^{INFx})_{1,\dots,x}, a_{1,t}^{SUP}, a_{1,t}^{INFx}, (a_{n(x),t}^{INFx})_{1,\dots,x-1,x+1,\dots,x} \right) \\ & \geq E \left( r_{\text{final}(t)}^{SUP} \middle| b_{x,t}=0, (b_{x,t})_{1,\dots,x-1,x+1,\dots,x} s_{i,t}^{SUP}, (s_{j(x),t}^{INFx})_{1,\dots,x}, a_{1,t}^{SUP}, a_{2,t}^{INFx}, (a_{n(x),t}^{INFx})_{1,\dots,x-1,x+1,\dots,x} \right) \end{aligned} \tag{23}$$

Solving this inequality results in

$$b_{x,N-1} \leq \frac{2c_{x,N-1} \times \left( \rho_{1,N-1}^{SUP} - \rho_{2,N-1}^{SUP} \right) \times \left( \alpha_{1.1.1,N-1}^{INFx} + \alpha_{1.2,N-1}^{INFx} - 1 \right)}{\left( \rho_{1,N-1}^{SUP} - \rho_{2,N-1}^{SUP} \right) \times \left( c_{x,N-1} \times \left( 2\alpha_{1.1,N-1}^{INF} - 1 \right) + \alpha_{1.1,SUP}^{SUP} \right) + \rho_{2,N-1}^{SUP}} \tag{24}$$

An incentive level that induces cooperation only exists if  $b_{x,N-1}^*$  satisfies (24). In the final section of this paper, we will illustrate this participation condition with a numerical example. Next, we determine the share coefficients for the earlier periods.

**Theorem 2** *The cost-minimal share coefficients  $b_{x,t}$  for  $t = 1, \dots, N - 2$  that motivate agents INFx to choose the cooperative action  $a_{1,t}^{INFx}$  only depends on data from the current period  $t$  and the next period  $t + 1$ . The share coefficient is*

$$b_{x,t}^* = \frac{\rho_{2,t}^{INFx} - \rho_{1,t}^{INFx}}{2c_{x,t} \left( \rho_{1,t}^{SUP} - \rho_{2,t}^{SUP} \right)} - \frac{\left( \alpha_{1.1,t+1}^{SUP} - \alpha_{2.1,t+1}^{SUP} \right) \times \left( \rho_{2,t+1}^{INFx} - \rho_{1,t+1}^{INFx} \right)}{2c_{x,t+1} \left( \rho_{1,t}^{SUP} - \rho_{2,t}^{SUP} \right)} \tag{25}$$

for  $t = 1, \dots, N - 2$

*Proof* see Appendix.

Theorem 2 shows that agents have to look merely one period ahead to determine their optimal decisions. Surprisingly, periods further into the future do not affect the current decisions. This fact can be explained by the anticipated choices of optimal share coefficients  $b_{x,t}^*$  in future periods. The optimal share coefficients make agents INFx indifferent between both of their actions in period  $t$ , i.e., the expected rewards are the same. Consequently, rewards of periods beyond the next period do not play a role in the decision process at the current epoch. For share coefficients  $b_{x,t}^* \neq b_{x,t}$  this is not the case, and data from all future periods are decision-relevant. We will show this effect in the next section, where agent SUP selects one share coefficient value for the entire time horizon.

#### 4.2 Multi-time-scale decision-making

In the multi-time-scale case, agent SUP commits to one type of action and one share coefficient value that applies to all periods. This case resembles a strategic decision. Agent SUP does not have the flexibility to change its initial decision. Another way to interpret this situation is that agent SUP makes one decision that controls the Markov chain over subsequent periods. In our example, the manager either calls or sends a letter, which affects the final state specifying contract renewal or no renewal.

The time-invariant share coefficient is denoted by  $b_{x,1..N-1}$  and the time-invariant action by  $a_{m,1..N-1}^{SUP}$  with the subscript indicating that they apply to periods 1 through  $N - 1$ . The multi-time-scale aspect does not affect agent SUP’s preferred decision, which is still action with index 1, i.e.,  $a_{1,1..N-1}^{SUP}$ .

To determine the optimal share coefficients, we first calculate intermediate, auxiliary share coefficient values in each period. We introduce an auxiliary share coefficient  $b_{x,[t]}$ , which is the basis for selecting the optimal value  $b_{x,1..N-1}^*$  for each agent INFx. Agent SUP then chooses the largest share  $b_{x,[t]}$  among the auxiliary share coefficients over the time horizon to ensure that each agent INFx chooses the cooperative action in every period, i.e.,

$$b_{x,1..N-1}^* = \max_{t=1,\dots,N-1} [b_{x,[t]}] \tag{26}$$

For the following theorem, we introduce a shorthand for the reward and transition probability differences, which simplifies the representation of results:

$$\Delta\rho_t^{SUP} = \rho_{1,t}^{SUP} - \rho_{2,t}^{SUP}, \quad \Delta\rho_t^{INFx} = \rho_{2,t}^{INFx} - \rho_{1,t}^{INFx} \tag{27}$$

$$\Delta\alpha_t^{SUP} = \alpha_{1,1,t}^{SUP} - \alpha_{2,1,t}^{SUP}, \quad \Delta\alpha_t^{INFx} = \alpha_{1,1,t}^{INFx} - \alpha_{2,1,t}^{INFx} \tag{28}$$

**Theorem 3** *The auxiliary share coefficient  $b_{x,[t]}$ , which is required in (26) to determine the optimal share coefficient  $b_{x,1..N-1}^*$ , can be calculated as follows:*

$$b_{x,[t]} = \frac{1}{2} \times \frac{\text{num}_{x,t}}{\text{den}_{x,t}} \quad \text{for } t = 1, \dots, N - 1 \tag{29}$$

where numerator ( $\text{num}_{x,t}$ ) and denominator ( $\text{den}_{x,t}$ ) are computed recursively as follows:

$$\text{num}_{x,N-1} = \Delta\rho_{N-1}^{\text{INF}x} \tag{30}$$

$$\text{num}_{x,t} = \Delta\alpha_{t+1}^{\text{INF}x} \times \text{num}_{t+1} + \Delta\rho_t^{\text{INF}x} \quad \text{for } t = 1, \dots, N - 2 \tag{31}$$

and

$$\text{den}_{x,N-1} = \Delta\rho_{N-1}^{\text{SUP}} \times c_{x,N-1} \tag{32}$$

$$\begin{aligned} \text{den}_{x,t} &= \Delta\alpha_{t+1}^{\text{INF}x} \times \text{den}_{x,t+1} + c_{x,t} \times \left( \Delta\rho_t^{\text{SUP}} + \Delta\rho_{t+1}^{\text{SUP}} \times \Delta\alpha_{t+1}^{\text{INF}x} + \Delta\rho_{t+2}^{\text{SUP}} \times \Delta\alpha_{t+2}^{\text{INF}x} \right. \\ &\quad \left. \times \Delta\alpha_{t+1}^{\text{INF}x} + \dots + \Delta\rho_N^{\text{SUP}} \times \Delta\alpha_N^{\text{INF}x} \times \Delta\alpha_{N-1}^{\text{INF}x} \times \dots \times \Delta\alpha_{t+1}^{\text{INF}x} \right) \\ &= \Delta\alpha_{t+1}^{\text{INF}x} \times \text{den}_{x,t+1} + c_{x,t} \times \left[ \Delta\rho_t^{\text{SUP}} + \sum_{\tau=t+1}^{N-1} \left( \Delta\rho_{\tau}^{\text{SUP}} \times \prod_{T=t}^{\tau} \Delta\alpha_{T+1}^{\text{INF}x} \right) \right] \\ &\quad \text{for } t = 1, \dots, N - 2 \end{aligned} \tag{33}$$

*Proof* see [Appendix](#).

The results of Theorem 3 show that in order to make decisions in the multi-time-scale case, agents SUP and INFx need information from all future periods. In contrast, for the single-time-scale case (Theorem 2), agents INFx only needed information from the current and next period. Besides the effect of higher information needs, there is a second effect associated with the multi-time-scale model: agent SUP has to pay a larger share of its reward to incentivize agents INFx. In the single-time-scale case, agent SUP had more flexibility to adjust incentives, which allowed agent SUP to retain more reward for itself.

In the final step, we determine if agent SUP wants to offer the optimal share coefficients  $b_{x,1..N-1}^* > 0$  that result in cooperative behavior by agents INFx, or rather not pay an incentive. Not paying an incentive allows agent SUP to retain all of its rewards, but results in non-cooperative behavior by agents INFx. Agent SUP commits to share coefficients  $b_{x,1..N-1}^* > 0$  according to (26) and Theorem 3, if its participation condition is met. Agent SUP decides on the level of incentive in the first period and informs agents INFx of its selection of the share coefficients. The participation condition can be determined as a closed-form analytic result similar to (24), but the size of the equation is too large to present effectively. Still, the participation condition can be easily calculated and evaluated for specific data using mathematical software. The following section provides a numerical example.

### 5 Numerical example

To provide further insights and apply the results of the model, we analyze a numerical example. The example illustrates the difference between the single-time-scale and the multi-time-scale model. In addition, we compare the expected rewards of cooperative infimal agents to the no-cooperation case, where agent SUP does not offer incentives and agents INFx choose non-cooperative actions.



**Table 1** Numerical results of analysis

	Single-time-scale	Multi-time-scale	No cooperation
	$b_{1,1} = 9.70\%$	$b_{1,[1]} = 9.75\%$	$b_{1,1} = 0$
	$b_{1,2} = 9.70\%$	$b_{1,[2]} = 9.89\%$	$b_{1,2} = 0$
	$b_{1,3} = 9.70\%$	$b_{1,[3]} = 10.39\%$	$b_{1,3} = 0$
	$b_{1,4} = 12.12\%$	$b_{1,1..5}^* = b_{1,[4]} = 12.12\%$	$b_{1,4} = 0$
$E\left(r_{\text{final}(1)}^{\text{SUP}}\right)$	187.53187.53	183.71	172.94
$E\left(r_{\text{final}(1)}^{\text{INF}x}\right)$	27.39	31.21	9.63

The example is based on the following time-invariant data of a four-period interaction ( $N = 5$ ) between agent SUP and one infimal agent.

$$R_t^{\text{SUP}} = \begin{pmatrix} 60 \\ 5 \end{pmatrix}, \quad R_t^{\text{INF}1} = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

$$P_t^y\left(s_{1,t}^y\right) = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}, \quad P_t^y\left(s_{2,t}^y\right) = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

$$C_t = c_{1,t} = 0.15$$

for  $t = 1, 2, 3, 4, y \in \{SUP, INF1\}$ . We assume that agents are in states with index 1 in the first period, i.e.,  $s_{1,1}^{\text{SUP}}$  and  $s_{1,1}^{\text{INF}1}$ . Table 1 shows the results of the analysis.

A number of observations can be made based on these results. First, agent SUP is better off choosing the optimal, non-zero share coefficients in both the multi-time-scale and single-time-scale case. These incentives induce cooperative behavior by agent INF1 since its expected reward is higher than if it did not cooperate. In other words, the participation condition is met for the chosen data. Furthermore, agent INF1 significantly benefits from the incentive and its cooperative actions, as its cumulative expected reward about triples from 9.63 to 27.39 and 31.21, respectively.

The second observation is that in the single-time-scale case, the share coefficient values are identical for periods 1, 2, and 3. The values are identical since the data of all three periods is the same. The share coefficient in the fourth and last period is larger than the prior ones. This deviation can be explained as follows: in earlier periods, agent INF1 benefited from transitioning with a higher probability to its preferred state. Thus, agent INF1 received a benefit beyond the current period’s reward, which allowed for a lower share coefficient as compensation. In the last period, this effect no longer exists, hence the larger share coefficient of 12.12 % compared to 9.70 %.

The third observation is that in the multi-time-scale model, the auxiliary share coefficient values differ in each period. The last period requires the largest share coefficient value of 12.12 %, which is applied to all periods. With data that is different for different periods, any period, not just the last, could have been the period with the highest auxiliary share coefficient.

Extending the example from one to multiple infimal agents yields similar results, since agents INF $x$ , their actions, and incentives are independent from one another. However, the cumulative influence  $C_t$  must not violate condition (10), which implies that individual change coefficients  $c_{x,t}$  will have to be smaller for a larger number of infimal agents. The consequence of smaller change coefficients is that incentives for agents INF $x$  need to increase to motivate cooperation, while the upper limit agent SUP is willing to pay for incentives (participation condition) goes down. Thus, the likelihood of cooperation decreases if rewards remain the same.

We conducted a sensitivity analysis with respect to the agents' rewards. We found that reward difference  $\Delta\rho_t^{\text{SUP}}$  has to be multiple times larger than  $\Delta\rho_t^{\text{INF}x}$  to result in situations in which agent SUP benefits from paying the necessary incentives that motivate cooperative behavior from agents INF $x$ . This effect intensifies as the number of infimal agents increases. This means that incentives will only be offered when the difference in payoffs between preferred and not preferred outcomes for the superior is sufficiently larger than that of their subordinates. For example, for  $\Delta\rho_t^{\text{SUP}} \leq 40$  instead of 55 (without any other changes) or for  $\Delta\rho_t^{\text{INF}x} \geq 5$  instead of 2, agent SUP prefers to not offer an incentive.

Furthermore, infimal agents' actions need to have a significant effect on agent SUP via change coefficient  $C_t$ . In the context of our example, this means that the financial benefits of customer contract renewal must be larger than the cost of providing customer care for the manager to offer incentives to her customer representatives. Furthermore, customer care must have a sufficiently strong influence on the chance of contract renewal. Otherwise, the manager will prefer not to pay an incentive and not to motivate cooperative behavior. For example, a reduction of the change coefficient to  $c_{1,t} \leq 0.12$  from 0.15 would result in no incentive payments.

The advantage of analytical solutions for sensitivity analysis is that one can readily see from equations which data influence the results and how. By differentiating the share coefficient equations and participation conditions with respect to model parameters, one can analytically describe the effect of changes in data. For an example of a comprehensive sensitivity analysis of multiscale decision making models see Wernz and Deshmukh (2007a) and Henry and Wernz (2013).

## 6 Discussion and conclusions

We developed a multi-time-scale decision-making model that allows hierarchically interacting agents to make optimal decisions over many periods. The supramal agent, agent SUP, chooses one type of decision and one level of incentive that is fixed over the entire decision horizon. If the participation condition is met, i.e., if profitable, agent SUP will offer a share of its pay (reward) to motivate cooperative actions by agents INF $x$ . We determined the optimal levels of incentives, which take into account multi-period agent interdependencies. In the context of the motivational example, this means that the account manager can determine a cost-minimal

incentive level that motivates cooperative behavior by customer representatives, which can resolve the representatives' conflicts of interest.

Furthermore, we showed the difference between a multi-time-scale model and a single-time scale model. In the single-time scale case, agent SUP (the manager) can change its decisions and level of incentives from period to period. The advantage of this flexibility is that the optimal incentive levels and decisions can be determined with a subset of data from the current and next period. In contrast, in the multi-time-scale, where agent SUP makes a strategic decision that applies to all periods, it needs information from all periods to find an optimal decision. In addition, agent SUP has to pay more, i.e., offer higher incentive levels, to obtain cooperative behavior from its infimal agents. The comparison of the single- and multi-time-scale models shows the cost of strategic management and the benefits of flexibility in terms of rewards and data requirements.

The paper contributes to the unification of the temporal and organizational scale in multi-agent, multi-period models. Furthermore, the paper provides an alternative formulation for multi-time-scale MDPs. Most results in this paper can be described analytically as opposed to being described only numerically. The advantage of analytic solutions is that agents know which information they need in order to make optimal decisions, the effect of parameters and the sensitivity of results can be readily determined, and optimal decisions can be calculated with little computational effort.

In the description of the example problem, we listed a number of questions that can now be answered.

1. *How high does the incentive for the representatives have to be to switch from the initially preferred action of being strict to the cooperative action of being lenient?* Theorems 1–3 answer this question by providing equations that calculate the minimum level of incentive necessary to motivate cooperative action. In the numerical example, the incentive that induces cooperative action from customer representatives, expressed in terms of the share of the manager's reward, was 9.7 and 12.12 %, respectively.

2. *What is the maximum share of her pay the manager is willing to offer before the cost of the incentives exceeds the benefits?* The participation condition formulated in (23) answers this question. For the single-time-scale case, equation (24) is the analytic solution for the maximum incentive level in the last decision epoch. For earlier periods, and for the multi-time-scale case, results can be obtained for specific data, as shown in the numerical example. The sensitivity analysis showed that for changes in parameters, such as  $\Delta\rho_t^{\text{SUP}} \leq 40$ ,  $\Delta\rho_t^{\text{INFx}} \geq 5$ , or  $c_{1,t} \leq 0.12$ , the participation condition is no longer met.

With an increasing number of customer representatives (agents INFx) that each individually exert a smaller influence on customers' decisions to renew their contracts, the manager's propensity to offer incentives goes down. The managerial implication of this insight is that incentives should be paid to small teams that effectively support their superiors' work. Once the team grows larger and the individual's contribution becomes smaller, incentives are no longer cost-effective for superiors.

3. *What role do the transition probabilities that describe the link between actions and outcomes play? How can the manager and the representatives compute their optimal decisions? What information do they need to determine an optimal decision?* The transition probabilities in the single-time-scale case play only a limited role in determining optimal incentives. For the incentives in the final period, transition probabilities do not affect the result (Theorem 1). For earlier periods, only the probabilities at the manager level (agent SUP) are needed to determine the optimal incentive levels (Theorem 2). This implies that customer representatives (agents INF $x$ ) can determine their optimal actions using partial or incomplete information. In particular, the incentive payments are independent of the action-outcome uncertainties (transition probabilities) at the customer representatives' level. In contrast, the manager needs data on all model parameters to determine whether she wants to pay the optimal levels of incentives. Furthermore, all data are needed by all agents in the multi-time-scale case. However, aggregated data in form of reward and probability differences are sufficient; see (27)–(28).

4. *How do the temporal dynamics of representatives interacting multiple times with a customer before contract renewal affect all of these aspects?* The difference in decision frequency is captured by the multi-time-scale model. When compared to the single-time-scale model, incentives for customer representatives have to be larger in the multi-time-scale case. This results primarily from the inflexibility of the manager to adjust the incentives from period to period. Consequently, customer representatives receive higher incentives and the manager has to share a higher percentage of her reward with her subordinates.

The multi-time-scale model considered only one decision horizon at the level of agent SUP (manager). An expansion to multiple decision horizons, i.e., a long-term perspective on interactions of manager, representatives and customers, is possible. Each decision horizon is coupled to the next only by the factor of whether a customer has renewed their contract or not. The analysis can thus be broken down into individual decision horizons and merely the initial states would differ. To explore a long or infinite decision horizon, the limit behavior of model parameters on optimal share coefficients would provide insights. Without much analysis, one can see that the effect of far-out periods on current periods is small, getting smaller with more temporal distance.

Future work should explore the effect of more hierarchical levels, more outcome states, and more actions. Wernz and Deshmukh (2012) modeled a 3-level, 3-period model, which illustrates the complexity of adding just one level in a multi-time-scale case. Still, a multi-level expansion that uses an algorithmic rule to capture all possible permutations and automatically analyzes all interaction scenarios seems possible. Sudhaakar and Wernz (2013) explored the effect of multiple actions and outcomes for a multi-level, one-period model. They showed that with more actions and outcomes, the transition probabilities, which previously were not necessary, are now needed to calculate the incentive levels. In a multi-period extension of this work, we expect to lose the ability to present the results as analytic solutions, but solutions based on data using analytical methods would still be possible.

**Acknowledgments** The author thanks the editor-in-chief Ahti Salo, the anonymous editor and the two anonymous referees for handling this paper and for providing constructive comments and improvement suggestions. This research has been funded in part by NSF grant CMMI-1335407 and by the Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) grant JFC-11-130.

### Appendix

*Proof for Theorem 1* We begin the analysis for a model with two infimal agents INF1 and INF2 and then show the generalization to  $X$  agents. First, we determine the share coefficient  $b_{1,N-1}$  that ensures cooperative behavior from agent INF1. To this end, we evaluate

$$\begin{aligned}
 & E\left(r_{\text{final}(N-1)}^{\text{INF1}} \mid s_{i,N-1}^{\text{SUP}}, s_{j(1),N-1}^{\text{INF1}}, s_{j(2),N-1}^{\text{INF2}}, a_{1,N-1}^{\text{SUP}}, a_{1,N-1}^{\text{INF1}}, a_{1,N-1}^{\text{INF2}}\right) \\
 & \geq E\left(r_{\text{final}(N-1)}^{\text{INF1}} \mid s_{i,N-1}^{\text{SUP}}, s_{j(1),N-1}^{\text{INF1}}, s_{j(2),N-1}^{\text{INF2}}, a_{1,N-1}^{\text{SUP}}, a_{2,N-1}^{\text{INF1}}, a_{1,N-1}^{\text{INF2}}\right)
 \end{aligned} \tag{34}$$

and solve the inequality for the share coefficients  $b_{x,N-1}$ . According to (19) and since  $E^*\left(r_{\text{final}(N)}^{\text{INF}x} \mid s_{k,N}^{\text{SUP}}, (s_{l(x),N}^{\text{INF}x})_{1,\dots,X}\right) = 0$ , inequality (34) is equivalent to

$$E\left(r_{\text{final},N-1}^{\text{INF1}} \mid \dots, a_{1,N-1}^{\text{INF1}}, \dots\right) \geq E\left(r_{\text{final},N-1}^{\text{INF1}} \mid \dots, a_{2,N-1}^{\text{INF1}}, \dots\right) \tag{35}$$

Applying the model’s parameters and coefficients, the left hand side of inequality (35) becomes

$$\begin{aligned}
 & \left(\rho_{1,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{1,N-1}^{\text{SUP}}\right) \times \alpha_{j(1),1,N-1}^{\text{INF1}} \times \alpha_{j(2),1,N-1}^{\text{INF2}} \times \left(\alpha_{i,1,N-1}^{\text{SUP}} + c_{1,N-1} + c_{2,N-1}\right) \\
 & + \left(\rho_{2,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{1,N-1}^{\text{SUP}}\right) \times \left(1 - \alpha_{j(1),1,N-1}^{\text{INF1}}\right) \times \alpha_{j(2),1,N-1}^{\text{INF2}} \\
 & \times \left(\alpha_{i,1,N-1}^{\text{SUP}} - c_{1,N-1} + c_{2,N-1}\right) + \left(\rho_{1,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{2,N-1}^{\text{SUP}}\right) \times \alpha_{j(1),1,N-1}^{\text{INF1}} \\
 & \times \alpha_{j(2),1,N-1}^{\text{INF2}} \times \left(1 - \alpha_{i,1,N-1}^{\text{SUP}} - c_{1,N-1} - c_{2,N-1}\right) + \left(\rho_{2,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{2,N-1}^{\text{SUP}}\right) \\
 & \times \left(1 - \alpha_{j(1),1,N-1}^{\text{INF1}}\right) \times \alpha_{j(2),1,N-1}^{\text{INF2}} \times \left(1 - \alpha_{i,1,N-1}^{\text{SUP}} + c_{1,N-1} - c_{2,N-1}\right) \\
 & + \left(\rho_{1,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{1,N-1}^{\text{SUP}}\right) \times \alpha_{j(1),1,N-1}^{\text{INF1}} \times \left(1 - \alpha_{j(2),1,N-1}^{\text{INF2}}\right) \\
 & \times \left(\alpha_{i,1,N-1}^{\text{SUP}} + c_{1,N-1} - c_{2,N-1}\right) + \left(\rho_{2,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{1,N-1}^{\text{SUP}}\right) \\
 & \times \left(1 - \alpha_{j(1),1,N-1}^{\text{INF1}}\right) \times \left(1 - \alpha_{j(2),1,N-1}^{\text{INF2}}\right) \times \left(\alpha_{i,1,N-1}^{\text{SUP}} - c_{1,N-1} - c_{2,N-1}\right) \\
 & + \left(\rho_{1,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{2,N-1}^{\text{SUP}}\right) \times \alpha_{j(1),1,N-1}^{\text{INF1}} \times \left(1 - \alpha_{j(2),1,N-1}^{\text{INF2}}\right) \\
 & \times \left(1 - \alpha_{i,1,N-1}^{\text{SUP}} - c_{1,N-1} + c_{2,N-1}\right) + \left(\rho_{2,N-1}^{\text{INF1}} + b_{1,N-1} \times \rho_{2,N-1}^{\text{SUP}}\right) \\
 & \times \left(1 - \alpha_{j(1),1,N-1}^{\text{INF1}}\right) \times \left(1 - \alpha_{j(2),1,N-1}^{\text{INF2}}\right) \times \left(1 - \alpha_{i,1,N-1}^{\text{SUP}} + c_{1,N-1} + c_{2,N-1}\right)
 \end{aligned} \tag{36}$$

The right hand side of (35) is similar to term (36) and the only difference is that  $\alpha_{j(1),1,N-1}^{INF1}$  is replaced by  $(1 - \alpha_{j(1),2,N-1}^{INF1})$ , and that  $(1 - \alpha_{j(1),1,N-1}^{INF1})$  by  $\alpha_{j(1),2,N-1}^{INF1}$ . Solving inequality (34) results in

$$b_{1,N-1} \geq \frac{\rho_{2,N-1}^{INF1} - \rho_{1,N-1}^{INF1}}{2c_{1,N-1}(\rho_{1,N-1}^{SUP} - \rho_{2,N-1}^{SUP})}. \tag{37}$$

Next, we verify that the decision of agent INF2, which was set at  $a_{1,N-1}^{INF2}$  in inequality (34), has no effect on result (37). With the alternative action  $a_{2,N-1}^{INF2}$  on both sides of inequality (34),  $\alpha_{j(2),1,N-1}^{INF2}$  is replaced by  $(1 - \alpha_{j(2),2,N-1}^{INF2})$ , and  $(1 - \alpha_{j(2),1,N-1}^{INF2})$  by  $\alpha_{j(2),2,N-1}^{INF2}$  on the left (36) and right hand side. The terms once again cancel each other out and the result is structurally identical to (37). The reason why the share coefficient of agent INF1 is not influenced by agent INF2 is that inequality (37) does not depend on any transition probabilities, only reward differences and agent INF1’s change coefficient.

Taking the two observations together—(1) the result independence from other infimal agent, and (2) the structural equivalence of result for both infimal agents—we can conclude that (37) holds for multiple infimal agents and for any agent INF*x*.

In the final step, we show why the incentive is at the lower bound of inequality (37). Since agent SUP’s reward decreases with a larger share coefficient, it will choose the smallest possible share coefficient that ensures cooperation. We assume that if an agent INF*x* is indifferent between the cooperative and non-cooperative action, it will choose the cooperative action. Consequently, agent SUP will choose the share coefficient such that (37) holds as an equality. In summary, we showed that solving

$$\begin{aligned} E\left(r_{\text{final}(N-1)}^{\text{INF}x} | s_{i,N-1}^{\text{SUP}}, \left(s_{j(x),N-1}^{\text{INF}x}\right)_{1,\dots,X}, a_{1,N-1}^{\text{SUP}}, a_{1,N-1}^{\text{INF}x}, \left(a_{n(x),N-1}\right)_{1,\dots,x-1,x+1,\dots,X}\right) \\ = E\left(r_{\text{final}(N-1)}^{\text{INF}x} | s_{i,N-1}^{\text{SUP}}, \left(s_{j(x),N-1}^{\text{INF}x}\right)_{1,\dots,X}, a_{1,N-1}^{\text{SUP}}, a_{2,N-1}^{\text{INF}x}, \left(a_{n(x),N-1}\right)_{1,\dots,x-1,x+1,\dots,X}\right) \end{aligned} \tag{38}$$

results in (22). □

*Proof for Theorem 2* We continue the analysis of Theorem 1 for  $N - 2$  and seek the share coefficient  $b_{x,N-2}$  that satisfies

$$\begin{aligned} E\left(r_{\text{final}(N-2)}^{\text{INF}x} | s_{i,N-2}^{\text{SUP}}, \left(s_{j(x),N-2}^{\text{INF}x}\right)_{1,\dots,X}, a_{1,N-2}^{\text{SUP}}, a_{1,N-2}^{\text{INF}x}, \left(a_{n(x),N-2}\right)_{1,\dots,x-1,x+1,\dots,X}\right) \\ = E\left(r_{\text{final}(N-2)}^{\text{INF}x} | s_{i,N-2}^{\text{SUP}}, \left(s_{j(x),N-2}^{\text{INF}x}\right)_{1,\dots,X}, a_{1,N-2}^{\text{SUP}}, a_{2,N-2}^{\text{INF}x}, \left(a_{n(x),N-2}\right)_{1,\dots,x-1,x+1,\dots,X}\right) \end{aligned} \tag{39}$$

The expression on the left hand side can be expressed as

$$\begin{aligned}
 & E\left(r_{\text{final},N-2}^{\text{INF}x} | s_{t,N-2}^{\text{SUP}}, \left(s_{j(x),N-2}^{\text{INF}x}\right)_{1,\dots,X}, a_{1,N-2}^{\text{SUP}}, a_{1,N-2}^{\text{INF}x}, \left(a_{n(x),N-2}^{\text{INF}x}\right)_{1,\dots,x-1,x+1,\dots,x}\right) \\
 & + \sum_{k=1}^2 \sum_{l(1)=1}^2 \dots \sum_{l(x)=1}^2 \dots \sum_{l(X)=1}^2 p_{\text{final},N-2}^{\text{SUP}} \left(s_{k,N-1}^{\text{SUP}} | s_{t,N-2}^{\text{SUP}}, a_{m,N-2}^{\text{SUP}}, \left(s_{l(x),N-1}^{\text{INF}x}\right)_{1,\dots,X}\right) \\
 & \times \prod_{x=1,\dots,X} p_{n-2}^{\text{INF}x} \left(s_{l(x),N-1}^{\text{INF}x} | s_{j(x),N-2}^{\text{INF}x}, a_{n(x),N-2}^{\text{INF}x}\right) \\
 & \times E^* \left(r_{\text{final}(N-1)}^{\text{INF}x} | s_{k,N-1}^{\text{SUP}}, \left(s_{l(x),N-1}^{\text{INF}x}\right)_{1,\dots,X}\right)
 \end{aligned} \tag{40}$$

We can regroup the terms of equation (39) to

$$\begin{aligned}
 & E\left(r_{\text{final}(N-2)}^{\text{INF}x} | \dots, a_{1,N-2}^{\text{INF}x}, \dots\right) - E\left(r_{\text{final}(N-2)}^{\text{INF}x} | \dots, a_{2,N-2}^{\text{INF}x}, \dots\right) + \sum_{k=1}^2 \sum_{l(1)=1}^2 \dots \sum_{l(x)=1}^2 \dots \\
 & \sum_{l(X)=1}^2 \left[ p_{\text{final},N-2}^{\text{SUP}}(\times) \times \left(p_t^{\text{INF}x}(\dots, a_{1,N-2}^{\text{INF}x}, \dots) - p_t^{\text{INF}x}(\dots, a_{2,N-2}^{\text{INF}x}, \dots)\right) \times E^*(\times) \right] = 0
 \end{aligned} \tag{41}$$

which simplifies to

$$\begin{aligned}
 & 2c_{x,N-2} \times b_{x,N-2} \times \left(\rho_{1,N-2}^{\text{SUP}} - \rho_{2,N-2}^{\text{SUP}}\right) - \left(\rho_{2,N-2}^{\text{INF}x} - \rho_{1,N-2}^{\text{INF}x}\right) \\
 & + \frac{c_{x,N-2} \times \left(\alpha_{1.1,N-1}^{\text{SUP}} - \alpha_{2.1,N-1}^{\text{SUP}}\right) \times \left(\rho_{2,N-1}^{\text{INF}x} - \rho_{1,N-1}^{\text{INF}x}\right)}{c_{x,N-1}} = 0
 \end{aligned} \tag{42}$$

and can be solved for the share coefficients resulting in

$$b_{x,N-2} = \frac{\rho_{2,N-2}^{\text{INF}x} - \rho_{1,N-2}^{\text{INF}x}}{2c_{x,N-2} \left(\rho_{1,N-2}^{\text{SUP}} - \rho_{2,N-2}^{\text{SUP}}\right)} - \frac{\left(\alpha_{1.1,N-1}^{\text{SUP}} - \alpha_{2.1,N-1}^{\text{SUP}}\right) \times \left(\rho_{2,N-1}^{\text{INF}x} - \rho_{1,N-1}^{\text{INF}x}\right)}{2c_{x,N-1} \left(\rho_{1,N-2}^{\text{SUP}} - \rho_{2,N-2}^{\text{SUP}}\right)}. \tag{43}$$

This result matches the theorem for  $t = N - 2$ . We continue with the backward induction and evaluate (39) for period  $N - 3$ . This yields the same result as in (43), with the time indices reduced by 1. Due to the self-similar structure of the equation over time, we can conclude that the result holds for every period  $t = 1, \dots, N - 2$ . □

*Proof for Theorem 3* We carry out the same analysis steps as in Theorems 1 and 2. We begin in decision epoch  $N - 1$ , and according to Theorem 1

$$b_{x,[N-1]} = \frac{\rho_{2,N-1}^{\text{INF}x} - \rho_{1,N-1}^{\text{INF}x}}{2c_{x,N-1} \times \left(\rho_{1,N-1}^{\text{SUP}} - \rho_{2,N-1}^{\text{SUP}}\right)}. \tag{44}$$

However, as the analysis continues for earlier periods, the close-form solutions become more complex. For  $t = N - 2$ , the result is

$$b_{x,[N-2]} = \frac{1}{2} \times \frac{\Delta\rho_{N-1}^{\text{INFx}} \times \Delta\alpha_{N-1}^{\text{INFx}} + \Delta\rho_{N-2}^{\text{INFx}}}{c_{x,N-1} \times \Delta\rho_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-1}^{\text{INFx}} + c_{N-2} \times (\Delta\rho_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-1}^{\text{SUP}} + \Delta\rho_{N-2}^{\text{SUP}})} \quad (45)$$

For  $t = N - 3$ , the transition function continues to be different and further increases in size. We present numerator and denominator of the share coefficient separately:

$$b_{x,[N-3]} \geq \frac{1}{2} \times \frac{\text{num}_{x,N-3}}{\text{den}_{x,N-3}}, \quad (46)$$

with

$$\begin{aligned} \text{num}_{x,N-3} &= \Delta\rho_{N-1}^{\text{INFx}} \times \Delta\alpha_{N-1}^{\text{INFx}} \times \Delta\alpha_{N-2}^{\text{INFx}} + \Delta\rho_{N-2}^{\text{INFx}} \times \Delta\alpha_{N-2}^{\text{INFx}} + \Delta\rho_{N-3}^{\text{INFx}} \quad (47) \\ \text{den}_{x,N-3} &= c_{x,N-3} \times [\Delta\rho_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-2}^{\text{SUP}} + \Delta\rho_{N-2}^{\text{SUP}} \times \Delta\alpha_{N-2}^{\text{SUP}} + \Delta\rho_{N-3}^{\text{SUP}}] \\ &\quad + \Delta\alpha_{N-2}^{\text{INFx}} \times [c_{x,N-1} \times \Delta\rho_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-1}^{\text{INFx}} + c_{x,N-2} \times (\Delta\rho_{N-1}^{\text{SUP}} \times \Delta\alpha_{N-1}^{\text{SUP}} + \Delta\rho_{N-2}^{\text{SUP}})] \quad (48) \end{aligned}$$

A repeating pattern for earlier periods for numerator and denominator can be observed. The patterns can be generalized and can be recursively described according to (29)–(33).  $\square$

## References

- Bellman RE (1957) Dynamic programming. Princeton University Press, Princeton
- Besanko D (1985) Multi-period contracts between principal and agent with adverse selection. *Econ Lett* 17(1–2):33–37
- Bhatnagar S, Panigrahi JR (2006) Actor-critic algorithms for hierarchical markov decision processes. *Automatica* 42(4):637–644
- Bowling M, Veloso M (2000) An analysis of stochastic game theory for multiagent reinforcement learning. School of Computer Science, Carnegie Mellon University, Pennsylvania
- Chang HS (2004) A model for multi-timescaled sequential decision-making processes with adversary. *Math Comput Model Dyn Syst* 10(3–4):287–302. doi:10.1080/13873950412331335261
- Chang HS, Fard PJ, Marcus SI, Shayman M (2003) Multitime scale Markov decision processes. *IEEE Trans Autom Control* 48(6):976–987
- Delebecque F, Quadrat JP (1981) Optimal control of Markov chains admitting strong and weak interactions. *Automatica* 17(2):281–296
- Dolgov D, Durfee E (2004) Graphical models in local, asymmetric multi-agent Markov decision processes. In: Proceedings of the third international joint conference on autonomous agents and multiagent systems, vol 2, pp 956–963
- Filar JA, Vrieze K (1996) Competitive Markov decision processes. Springer, New York
- Goseva-Popstojanova K, Trivedi KS (2000) Stochastic modeling formalisms for dependability, performance and performability. In: Haring G, Lindemann C, Reiser M (eds) Performance evaluation—origins and directions., Lecture Notes in Computer ScienceSpringer, New York, pp 403–422
- Hauskrecht M, Meuleau N, Kaelbling LP, Dean T (1998) Boutilier C Hierarchical solution of Markov decision processes using macro-actions. In: Proceedings of the fourteenth conference on uncertainty



- in artificial intelligence. University of Wisconsin Business School, Madison, WI, July 24–26, pp 220–229
- Henry A, Wernz C (2013) Revenue-sharing in a three-stage supply chain with uncertainty: a multiscale decision theory approach (under review)
- Howard RA (1960) Dynamic programming and Markov process. MIT Press, Cambridge, MA
- Jacobson M, Shimkin N, Shwartz A (2003) Markov decision processes with slow scale periodic decisions. *Math Oper Res* 28(4):777–800
- Marschak J (1955) Elements for a theory of teams. *Manag Sci* 1(2):127–137
- Marschak J, Radner R (1972) Economic theory of teams. Yale University Press, New Haven
- Mertens JF (1992) Stochastic games. In: Aumann RJ, Hart S (eds) *Handbook of game theory with economic applications*. North-Holland, Amsterdam
- Mesarovic MD, Macko D, Takahara Y (1970) Theory of hierarchical, multilevel systems. Academic Press, New York
- Milgrom PR, Roberts J (1992) Economics organizations and management. Prentice Hall, Englewood Cliffs, NJ
- Muppala JK, Malhotra M, Trivedi KS (1996) Markov dependability models of complex systems: analysis techniques. In: Özekici S (ed) *Reliability and maintenance of complex systems*. Springer, Berlin, pp 442–486
- Panigrahi JR, Bhatnagar S (2004) Hierarchical decision making in semiconductor fabs using multi-time scale Markov decision processes. In: *Proceedings of IEEE conference on decision and control, Paradise Island, Nassau, Bahamas, 2004*. IEEE, pp 4387–4392
- Parr RE (1998) Hierarchical control and learning for Markov decision processes. PhD thesis, University of California, Berkeley
- Plambeck EL, Zenios SA (2000) Performance-based incentives in a dynamic principal-agent model. *Manuf Serv Oper Manag* 2(3):240–263
- Pollatschek MA, Avi-Itzhak B (1969) Algorithms for stochastic games with geometrical interpretation. *Manag Sci* 15(7):399–415
- Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York
- Radner R (1992) Hierarchy: the economics of managing. *J Econ Lit* 30(3):1382–1415
- Raghavan TES, Filar JA (1991) Algorithms for stochastic games—Wa survey. *Math Methods Oper Res (ZOR)* 35(6):437–472
- Savage LJ (1954) *The foundation of statistics*. Wiley, New York
- Schneeweiss C (1995) Hierarchical structures in organizations: a conceptual framework. *Eur J Oper Res* 86(1):4–31
- Schneeweiss C (2003a) *Distributed decision making*. Springer, Berlin
- Schneeweiss C (2003b) Distributed decision making—a unified approach. *Eur J Oper Res* 150(2):237–252
- Sethi SP, Zhang Q (1994) *Hierarchical decision making in stochastic manufacturing systems*. Birkhäuser, Basel
- Shapley LS (1953) Stochastic games. *Proc Natl Acad Sci* 39(10):1095–1100
- Simon HA (1957) *Models of man*. Part IV. Wiley, New York
- Simon HA (1972) Theories of bounded rationality. In: Mcguire C, Radner R (eds) *Decision and organization*. North-Holland, Amsterdam, pp 161–176
- Sudhaakar S, Wernz C (2013) Advancing multiscale decision theory: from 2 to  $N$  decision alternatives. Working Paper
- Sutton RS (1995) TD Models: modeling the world at a mixture of time scales. In: *Proceedings of the twelfth international conference on machine learning, Tahoe City, CA, July 9–12, 1995*, vol 95, pp 531–539
- Sutton RS, Precup D, Singh S (1999) Between MDPs and Semi-MDPs: a framework for temporal abstraction in reinforcement learning. *Artif Intell* 112(1–2):181–211
- Weber M (1978) *Economy and society*. University of California Press, Berkeley
- Wernz C (2008) Multiscale decision making: bridging temporal and organizational scales in hierarchical systems. University of Massachusetts Amherst, Dissertation
- Wernz C, Deshmukh A (2009) An incentive-based, multi-period decision model for hierarchical systems. In: *Proceedings of the 3rd annual conference of the Indian subcontinent decision sciences institute region (ISDSI), Hyderabad, India, 2009*

- Wernz C, Deshmukh A (2007a) Decision strategies and design of agent interactions in hierarchical manufacturing systems. *J Manuf Syst* 26(2):135–143
- Wernz C, Deshmukh A (2007b) Managing hierarchies in a flat world. In: Proceedings of the 2007 industrial engineering research conference. Nashville, TN, pp 1266–1271
- Wernz C, Deshmukh A (2010a) Multi-time-scale decision making for strategic agent interactions. In: Proceedings of the 2010 industrial engineering research conference, Cancun, Mexico
- Wernz C, Deshmukh A (2010b) Multiscale decision-making: bridging organizational scales in systems with distributed decision-makers. *Eur J Oper Res* 202(3):828–840
- Wernz C, Deshmukh A (2012) Unifying temporal and organizational scales in multiscale decision-making. *Eur J Oper Res* 223(3):739–751
- Wernz C, Henry A (2009) Multilevel coordination and decision-making in service operations. *Service Sci* 1(4):270–283
- Williamson OE (1967) Hierarchical control and optimum firm size. *J Political Econ* 75(2):123–138
- Williamson OE (1970) Corporate control and business behavior. Prentice Hall, Engelwood Cliffs, NJ
- Williamson OE (1979) Transaction-cost economics: the governance of contractual relations. *J Law Econ* 22(2):233–261. doi:[10.2307/725118](https://doi.org/10.2307/725118)
- Wongthatsaneorn W, Realff MJ, Ammons JC (2010) Multi-time scale Markov decision process approach to strategic network growth of reverse supply chains. *Omega* 38(1–2):20–32
- Yeow W-L, Tham C-K, Wong W-C A (2005) Novel target movement model and energy efficient target tracking in sensor networks. In: IEEE 61st Vehicular Technology Conference, 2005:2825–2829
- Yeow WL, Tham C-K, Wong W-C (2007) Energy efficient multiple target tracking in wireless sensor networks. *IEEE Trans Veh Technol* 56(2):918–928
- Yin G, Zhang Q (1998) Continuous-time Markov chains and applications: a singular perturbation approach. Springer, Berlin
- Yin G, Zhang Q (2005) Discrete-time Markov chains: two-time-scale methods and applications. Springer, Berlin
- Zachrisson LE (1964) Markov games' advances in game theory. *Ann Math Stud* 52:211–253
- Zhu C, Zhou J, Wu W, Mo L (2006) Hydropower portfolios management via markov decision process. In: IECON 2006—32nd annual conference on IEEE industrial electronics, 2006:2883–2888