Can multiple contractors self-regulate their joint service delivery? A serious gaming experiment on road maintenance planning

Joris Scharpff , Daan Schraven , Leentje Volker , Matthijs T. J. Spaan & Mathijs M. de Weerdt

Published online: 17 Aug 2020.

Submit your article to this journal ↗

Article views: 445

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Can multiple contractors self-regulate their joint service delivery? A serious gaming experiment on road maintenance planning

Joris Scharpff[a], Daan Schraven[b] , Leentje Volker[c] , Matthijs T. J. Spaan[a] and Mathijs M. de Weerdt[a]

[a]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands; [b]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands; [c]Faculty of Engineering Technology, University of Twente, Enschede, Netherlands

ABSTRACT

The next step in the use of innovative, dynamic and performance-based contracts for service delivery by contractors could be use of monetary incentives to stimulate self-regulation of the network. Because it is currently unclear how performance-based payments in network tenders can effectively encourage network members to coordinate their own operations, a serious game was performed that simulates road maintenance planning to study changes in decision making and the emergence of network coordination. The experiments show that monetary incentives influence decision making, but their effect may be opposite to their intended aim and can lead to a competitive network. It was, however, also found that this competitiveness is not shown in networks where members are familiar with each other. This leads to the conclusion that penalty-based incentive mechanisms probably interfere with self-regulation and that the social dimension of contractor collaboration is paramount to the success of network-based contracting of construction activities.

Introduction

Over the two last decades, performance-based contracting has become the prevalent approach for service delivery in public-private partnerships (Smith and Grinker 2004, Demirel *et al.* 2017). In these contractual agreements, the performance as contracted is often scoped by the contractors' specific project activities, e.g. a delay in time or a traffic jam on a particular road undertaken by construction activities. Hence, as public clients are slowly changing towards a stewardship role (Snippert *et al.* 2015, Potemans *et al.* 2018), the expectation rises for contractors to also perform well on effects that fall outside of their project scope, like construction activities causing a traffic jam elsewhere on the infrastructure network.

In practice, this kind of multi-lateral agreements could, for example, be organised on a set of regional roads, where multiple contractors together manage various adjacent areas covering an entire province. Hence, this leaves a vacuum regarding the expertise on the coordination of such activities. An important challenge for this modus operandi is to ensure that a group of competitive contractors perform optimally under such multilateral collaborative agreements (Bresnen and Marshall 2000, Rose and Manley 2011). Of particular interest is the use of incentives to not only maximise performance, but also instil other performance aspects, such as self-regulation (Agranoff and McGuire 2001). This approach is actively being pursued by, amongst others, Volker *et al.* (2012) and Hosseinian and Carmichael (2013). If applied well, self-regulation implements the key ideas of performance-based contracting, i.e. letting the contractors "do what they do best" and account them for their performance, while additionally stimulating them to coordinate their operations amongst themselves.

The design of monetary incentives is considered key to the success of these approaches. Since a perfectly designed payment mechanism is rendered useless if contractors are not stimulated by these incentives in their construction activities, of at least equal importance is the effectiveness of such incentives to influence the decision-making process of

CONTACT Joris Scharpff ✉ j.c.d.scharpff@tudelft.nl 🖂 Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, Netherlands

contractors. Although the use of incentives and requirements have received much attention in public network-management literature (Klijn *et al.* 2010a, Van Bueren *et al.* 2003), the effectiveness of such incentives has not been widely studied in the context of networks, let alone seen any empirical evaluation.

This study addresses the effectiveness of monetary incentives in network-based procurement strategies. In particular, it investigates the emergence of coordination in a setting that is designed to reward contractors for coordinating their interactions on a network level. This setting is created through serious gaming, a method to perform research in a simulated, closed environment. This approach enables controlled research into changes in behaviour or decision making without the cost or consequences associated with real-world testing (Meijer *et al.* 2012, Calderón and Ruiz 2015). Using such a serious game, the decision making of contractors is studied and their interactions are observed with the goal to validate the concept of self-regulation. That is, it is investigated whether contractors change their decision-making behaviour in the presence of network incentives and whether this leads to intrinsic self-regulated coordination of the network.

Furthermore, as several scholars report on the importance of relationships and their impact on contract outcomes, (Rose and Manley 2011, Demirel *et al.* 2017), the social cohesion of the network is investigated as a potential moderator for self-regulation. The change in behaviour due to incentives, the emergence of coordination and the role of social cohesion in self-regulation, are formulated as the three main research questions to be validated in this serious game. Ultimately, this validation forms a first proof of the concept of self-regulation of networks.

This study is organised as follows. The next section discusses the context of self-regulation, related work on the subject and formulates the research questions for this study. The following two sections introduce respectively the methodology and outline the serious game design and its gameplay. Thereafter, the results of the gaming sessions are presented and analysed. The paper closes with an in-depth discussion on the findings and their relevance.

Self-regulation in contracts

Contracting has recently seen a shift from traditional regulatory agreements towards performance-based partnerships and outsourced service delivery (Bresnen and Marshall 2000, Eriksson 2010). It is recognised widely that transferring autonomy from client to contractor offers many promising advantages over complete governance, fundamental to service-delivery partnerships. To name a few of these advantages: the ability to adapt to changing circumstances (Javed *et al.* 2014, Demirel *et al.* 2017), optimal use of contractor expertise, capabilities and means (Smith and Grinker 2004, Straub 2010), fair distribution of risks (Zietlow 2005), stimulation of innovation and efficiency in service-delivery (Zietlow 2005, Hughes and Kabiri 2013, Lam and Gale 2014), and the possibility of co-creation (Hartmann *et al.* 2014). Indeed, practical successes are reported in domains such as transportation infrastructure (Choi *et al.* 2011), road maintenance (Zietlow 2005, Lam and Gale 2014) and project construction (Love *et al.* 2011, Hughes and Kabiri 2013, Demirel *et al.* 2017).

Collaboration between multiple public and private parties, however, comes with additional complexities to that need to be managed by the contractual framework. Eriksson (2010) notes, for example, that cooperation and the benefits of partnering are not easily obtained due to a lack of understanding of the concept and when and how to implement it. Parkhe (1993) and Altamirano *et al.* (2008) highlight in particular the increased opportunity of opportunistic behaviour ("cheating") and additional uncertainty due to the greater complexity of the environment. The issue of opportunistic behaviour is also identified by Gao and Liu (2019), who identify this as a particular risk during the operational stage. With respect to complexity, Schraven *et al.* (2011) identifies the management of multiple actors with different interests as a main challenge, which is related to the conclusion by Shadid (2018) that a dynamic strategic framework is necessary to cope with the complexity of managing organisations in this setting. The social dimension also becomes more prevalent in networks. Bower *et al.* (2002), Van Bueren *et al.* (2003), Klijn *et al.* (2010a), Rose and Manley (2011) report that relationships, trust and alignment are key when dealing with a network. More strongly, Snippert *et al.* (2015) show that if not managed properly, performance-based approaches can result in a fall back to control-oriented relationship characterised by information asymmetry, lack of transparency and distrust.

Over the years the topic of innovative contracting has received a lot of attention in research which led to many requirements for their design. Demirel *et al.* (2017) and Javed *et al.* (2014) state that innovative contracts require a built-in flexibility do deal with contingencies, foreseen changes and complex interactions. Moreover, Kuitert *et al.* (2019) argue that this

flexibility is additionally required to integrate the interests of the network user into a contract. This is also suggested by Bower *et al.* (2002) who stress the importance of contractor participation in the design of incentives. Nalbantian and Schotter (1997) concluded that relative performance schemes outperform target-based schemes, thus supporting the idea of stimulating emergent coordination over individual performance objectives. Gupta *et al.* (2011) found that in a competitive bidding process, the additional cost of incentives is not larger than the additional value gained by higher-quality work. So, while these findings constitute substantial challenges, they concurrently illustrate the need for performance-based contracting solutions tailored to multilateral agreements.

The subject of inciting a network of multiple parties to regulate itself during execution, however, has not seen many contributions in literature either. The concept of self-regulation refers to the ability of a network to coordinate their interactions and dependencies without external regulation (Van Bueren *et al.* 2003). Almost 20 years ago, Agranoff and McGuire (2001) already raised the question "In what ways do networks employ mutual self-responsibility, and does this substitute for the loss of public agency accountability?" as one of the seven biggest open problems in network management. Since then, the potential of self-regulating networks in contracting frameworks is actively being investigated. Amongst others, Volker *et al.* (2014) highlight the promising benefits to all stakeholders of 1) using innovative, flexible contract forms based on incentives to engage and monitor service-delivery involving multiple contractors, and 2) the implementation and coordination of service-delivery using recurrent self-regulation by the network.

Although the notion that collaboration of a network could be "engineered" through incentive systems was ventured by Bresnen and Marshall (1998), the same authors surveyed that it has mostly led to a multitude of guidelines on partnering and alliances (Bresnen and Marshall 2000). Only a handful of works address the incentive design itself. Without exception, these approaches are based on monetary incentives that either reward collaboration in a project, penalise selfish behaviour or apply a combination of both. Best known are the revenue/cost sharing mechanisms typically referred to as respectively "gainshare" and "painshare" contracts, or a combination thereof (Bresnen and Marshall 2000, Hosseinian and Carmichael 2013). More specifically, Hosseinian and Carmichael (2013) consider the optimal design of gainshare/painshare incentives using the model of agency theory. This concept is extended to include the network user by Volker *et al.* (2012) who propose a sharing of social costs to stimulate contractor co-operation.

This novel idea has gotten form in the approach by Scharpff *et al.* (2013) who propose a game-theoretically designed incentive mechanism that incorporates the network user in its payments to stimulate coordination of maintenance activities and at the same time discourages opportunistic behaviour. Nonetheless, the absence of empirical evidence makes it is unclear whether such perfectly-engineered incentives achieve their intended goals when confronted with real actors (Volker *et al.* 2012). In this context, Rose and Manley (2011) and Turrini *et al.* (2010) raise a similar concern that despite an overall believe that incentive mechanisms improve value for money during procurement and project performance during execution, empirical research is scarce.

In parallel, researchers have raised several critical notes on the use of monetary incentives in contracting procedures to stimulate performance. Bresnen and Marshall (2000) stated, for example, that there are limitations to the use of incentives as means of reinforcing collaboration and developing commitment and trust, and that cognitive and social dimensions strongly affect the impact of incentives. A similar finding was also reported by Rose and Manley (2011) who conclude financial incentives to be less important to motivation and performance than relationship enhancement initiatives. According to the survey done by Turrini *et al.* (2010), network integration is a structural characteristic that emerges as a major determinant of network effectiveness, thereby referring to the positive of impact cohesion. Klijn *et al.* (2010a) and Volker *et al.* (2014) insist that in particular trust between network participants strongly affects the outcomes of the process and also Dewulf and Kadefors (2012) argue that incentives are important in shaping interaction and signalling trust.

Summarising the current state of the literature on performance-based contracting of groups, there is a consensus that the use of innovative, performance-based contracts in partnerships benefit all stakeholders, while at the same time there is much uncertainty with regards to the design of incentives and their effectiveness to incite self-regulation in actual tenders. This impasse is strengthened by the absence of any substantial experimental evidence or real-world successes. This thesis set out to provide a first proof of the concept of self-regulation in performance-based

contracts, provide guidance on its design and contribute techniques to optimise the value of such tenders. The next section introduces the main research question and its decomposition into several smaller challenges that are addressed by the chapters of this thesis.

> RQ_1 Are monetary incentives effective in influencing the decision making of contractors in the network setting?

> RQ_2 Does the use of monetary incentives that reward coordination of the network lead to self-regulation?

> RQ_3 Does the relationship between network members influence the effectiveness of the incentives and, indirectly, the manifestation of self-regulation?

Methodology

While ideally the research questions are answered through real-world experiments, this introduces additional and complex variables into the decision making and the cost of failure in practical tenders prohibits such an approach without prior evidence. Therefore, the concept of self-regulation though incentives should first be validated in a controlled, agent-based simulation in the form of a serious game. While serious games are primarily known as a tool for instruction, they can also be employed as a research method comparable to simulation or experimentation (Calderón and Ruiz 2015). Serious games have the potential to integrate a multiplicity of elements including motivation, expertise and social structures, and can be considered a means to explore, explain, and assess the complex interactions between ecosystems and human actions (Axelrod 2003). This research methodology is also known as (agent-based) participatory simulation (Guyot and Honiden 2006) and can serve to enhance our scientific understanding or to recommend corrective policy action, as in the studies of e.g. Le Bars and Le Grusse (2008), Altamirano et al. (2008) and Meijer et al. (2012). Although the closed system of a game does not allow direct generalisation to real-world tenders, it does enable a controlled evaluation of changes in the decision-making process and allows reflection on the concept of implementing self-regulation through incentives.

The analysis of incentives and changes in agent behaviour is performed within the framework of (agent-based) computational economics (see for instance (Tesfatsion and Judd 2006)), a field that lies at the interface of computer science, economics and management science. The approach in this paper follows the game-theoretical model of agent behaviour and decision making (Nisan et al. 2007), a model advocated by Parkhe (1993) for studies on agent behaviour and inter-firm cooperation or agency theory models (Eisenhardt 1985), and used in related studies by e.g. Roth (2002), Javed et al. (2014), Gao and Liu (2019). In essence, game theory models human decision makers as agents that play a game by performing actions that result in associated utilities. As a rule, it is assumed that agents are rational, or at least boundedly rational, and hence always choose actions that maximise their utility to the best of their knowledge and capability (Gigerenzer and Selten 2002, Kahneman 2003). This model can clearly illustrate the potential of monetary incentives: they change the utilities in the game, thereby steering agents to different action choices. How to design incentives to steer towards favourable outcomes is the main topic of mechanism design, see for example the work by Dash et al. (2003), and studied by Gupta et al. (2011) and Scharpff et al. (2013), but is not addressed here.

The context of the serious game is provided by an exemplary problem from the domain of infrastructural maintenance planning called the Road Maintenance Planning problem. The problem was first introduced through a game called "Road Roles" to study the effect of incentives on opportunistic behaviour by Altamirano et al. (2008) and later extended by Scharpff et al. (2013) to the problem of optimally planning maintenance operations by a network. In brief, it models a network of contractors ("service providers") that is responsible for the maintenance of a shared infrastructure. The essence of the game is that while each contractor plans its own maintenance work individually, they are accounted for their collective impact to the network user as a result of traffic hindrance on the network-level through monetary incentives. Put simply, concurrent maintenance leads to steep penalties that can only be avoided by coordinating operations on a network level, that is, by self-regulation of the network. Hence, by playing the game multiple times and observing the interactions both inside and outside the game, the goal is to study the effectiveness incentives to encourage self-regulation in a network of human decision makers and the influence of social cohesion.

### Operationalising decision-making, coordination and cohesion in game design

The decision preferences of players are expressed in terms of the three distinct decision parameters or

Question 5

You have chosen four potential periods in which you can perform your project. Using quarterly figures, you determine the following prospects regarding four possible maintenance periods:

Situation	Risk of delay		Period 1	Period 2	Period 3	Period 4
Project as planned	67%	TTL	370.000	416.000	333.000	615.000
		Profit	€ 1.369.000	€ 1.323.000	€ 1.406.000	€ 1.124.000
Project is delayed	""!	TTL	503.000	571.000	493.000	809.000
		Profit	€ 1.236.000	€ 1.168.000	€ 1.246.000	€ 930.000

In addition, you also possess information regarding the TTL figures of the previous year.
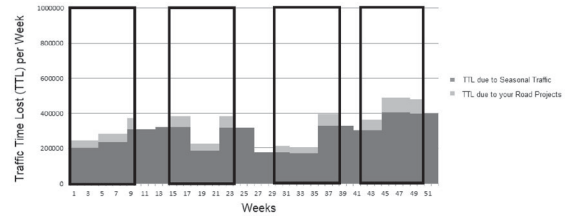
a) Can you specify the order of periods in which you prefer to perform the maintenance? Please rank them from 1 (best) to 4 (worst).

Answer:	Rank:
Period 1	
Period 2	
Period 3	
Period 4	

b) Please motivate your ranking.

Figure 1. Example problem from the questionnaire. Participants are given four alternative periods in which they can plan their work, each with different expected profits and traffic time losses that may occur in one of the two possible scenarios. They are asked to rank them according to their preference and provide a motivation for their ranking.

objectives of the problem domain, namely *profit*, *traffic time lost (ttl)*[1] and *risk aversion*. A preference for profit reveals itself in predominantly making choices that lead to a higher profit to the player. A player with a ttl preference is more likely to strive for minimisation of expected impact on the traffic hindrance. Players that are risk-averse tend to favour approaches with low delay probability, preferring more robust planning with low variance in both the profits and ttl.

The decision preferences of players before playing the game are captured in their *profile score*. This score is measured through a questionnaire (Figure 1) that participants submit 2 weeks prior to the game. This questionnaire consists of 8 increasingly complex game situations and participants are asked to rank the options according to their preferences. From this the profile scores are computed using the methodology from Triantaphyllou (2013). First the questionnaire is modelled as a multi-criteria decision making in which participants rank their preference over alternatives for every question to determine the relative weights of alternatives. Then, from the responses that specify rankings for the alternatives, a preference score per objective is computed using the weighted-sum method from Roszkowska (2013). Finally, the preference scores are normalised to obtain the relative importance of each objective. That is, if $\widehat{Q_p}(\mathbf{x})$ expresses the preference score for the objective profit (hence the subscript p) computed from questionnaire responses $\mathbf{x}$ as just described, the profile score for the profit parameter is computed by $Q_p(\mathbf{x}) = \widehat{Q_p}(\mathbf{x})/(\widehat{Q_p}(\mathbf{x}) + \widehat{Q_t}(\mathbf{x}) + \widehat{Q_r}(\mathbf{x}))$. Similarly, the profile scores $Q_t(\mathbf{x})$ and $Q_r(\mathbf{x})$ express the relative preference for respectively ttl (subscript t) and risk aversion (subscript r). Finally, the rationality of these responses is determined by comparing them to the

Pareto optimal trade-off (Rousis 2011), which serves as a measured for how well participants are able to make trade-offs.

The strategy scores are determined from the actions that are played in the game, roughly similar to the multi-criteria method used on the questionnaire. Each of the actions of the game have been designed for a specific goal and are assigned weights accordingly. Then the alternative weights of the actions performed are summed to form a preference score which in turn is normalised to express relative importance of the objective. For example, the "low hindrance" action will have only a minor impact on ttl and have a high ttl strategy score G_t. On the other hand, its costs are relatively high compared to the other available actions and thus the profit strategy score G_p for this action will be low. Given a maintenance plan plan $\mathbf{y}$ that describes the assignment of chosen maintenance methods to time slots, the profit strategy score for that agent is denoted by $G_p(\mathbf{y})$, using the same subscripts as before.

The outcomes of the game are measured in terms of the profit and ttl that are to be expected when executing the (current) plan $\mathbf{y}$, accounting for the risk of potential delays using the standard approach of probability times utility. The functions $P(\mathbf{y})$ and $T(\mathbf{y})$ express respectively the total expected profit and total expected traffic time lost, given joint plan $\mathbf{y}$. Additionally, the functions $P_w(\mathbf{y})$ and $T_w(\mathbf{y})$ represent respectively the profits and ttl in the worst case, e.g. when all maintenance operations delay. Note that the ttl depends on the joint plan of all players, i.e. all concurrent maintenance within the infrastructure. The utility to a player is defined as the sum of its profits from maintenance work minus its costs due to traffic penalties (1 euro for every hour of ttl in this game model).

Table 1. Definition of the qualitative categories for the coordination and cohesion level variables and their descriptions.

	Description
Coordination level	
Low	Conflict-driven coordination of interactions via bilateral or trilateral negotiations
Medium	Coordination of network via democratic, plenary negotiations
High	Centralised planning that governs network decisions
Cohesion level	
Unfamiliar	Players have (had) limited to no interaction previously
Familiar	Players see and/or work with each other on a regular basis

Thus the sum of utilities $u(\mathbf{y})$ for a plan $\mathbf{y}$ is given by $u(\mathbf{y}) = P(\mathbf{y}) - T(\mathbf{y})$, and similar for the worst case. Finally, an outcome of the game can be thought of as "better" if either the total profit increases, the total ttl decreases, or both. Therefore the *performance ratio* ϕ of a joint plan $\mathbf{y}$ is computed by the formula $\phi(\mathbf{y}) = P(\mathbf{y})/T(\mathbf{y})$. For clarity of presentation, the questionnaire response $\mathbf{x}$ or joint plan $\mathbf{y}$ are implicit in the notation, for instance $Q_t(\mathbf{x})$ and $P(\mathbf{y})$ are written as Q_t and P.

In addition to the quantitative information that is measured from the questionnaire and the game on decision preferences and the outcomes of the game, two qualitative parameters are considered in this study. These are the *coordination level* that is showcased by the network and the *social cohesion* of network members. Both parameter are established by classification rules, shown in Table 1. The coordination level of a network is established trough observing the communications and interactions that take place between the players outside of the game. Depending on the types of interaction displayed by the players, the game session is classified into one of the categories. For example, a session where players only resolve planning conflicts in pairs or triples would be categorised as "Low". If in addition plenary negotiations are used (at least once) it would be considered "Medium". The designation "High" is given to game sessions in which central regulation takes place. The social cohesion of the group is a qualification of the relation between participants outside of the game, established through inquiry at the start. Because a classification of social cohesion is hard to capture exactly, two clearly disjoint categories are used. If players in a gaming session have never or rarely interacted with each other, either during work or during social events, the group is classified as "Unfamiliar". Other groups, in which participants do frequently interact, are classified as "Familiar".

Hypothesising self-regulation in road networks

Using the methodology of the preceding sections, the research questions at the origin of this paper are examined. Therefore these questions are translated into hypotheses that can be tested in the closed system of the serious game. To ascertain that the game is designed correctly and the human decision makers act according to the model adopted here, first a set of validation hypotheses are formulated. Thereafter it is discussed how the original questions (albeit in a compacter form) can be answered through hypothesis testing.

Is the model valid?

As a first step, the model of the game itself is analysed to confirm the validity of the assumptions underlying the design and further experiments. To this end, three hypotheses are proffered. The first hypothesis is that the decision-making of human decision makers can be considered at least boundedly rational. While an exact degree of rationality cannot be determined, decision optimality can be used as an approximation of the ability to rationally maximise utility. Hence, the rationality of a questionnaire response, and thus an indicator for the rationality of the participant, is defined as its relative position on a scale from the lowest possible score to the Pareto-optimal trade-off that is closest to the response score. The resulting metric is an indicator for the rationality on the [0, 1] scale and agents are considered boundedly rational if this their rationality score is at least 0.8. Therefore it is hypothesised that the mean of rationality scores of the agents is at least 0.8 with a confidence level of 95%.

The second and third hypotheses establish the correctness of the game design. Validation hypothesis two states that the actions affect their designated objective. That is, the hypothesis is that there is a strong correlation between the played strategy scores G_p, G_t and G_r and their corresponding impact on the outcome $\mathbf{y}$, respectively the expected profit $P(\mathbf{y})$ and ttl $T(\mathbf{y})$ for the first two and the worst case profit $P_w(\mathbf{y})$ and ttl $T_w(\mathbf{y})$ for the latter. Finally, to confirm that coordination of decisions is beneficial to the players and hence the premise that network-based incentives stimulate coordination, a third hypothesis states

that the coordination level is strongly correlated to the utility.

Are monetary incentives effective to influence decision making?

It is hypothesised that applying monetary incentives in a network-based tender is an effective way to influence the decision-making process contractors. Hence there should be a notable difference in the decisions-making preferences of players when comparing the settings with and without monetary incentives. Although both the questionnaire and the game confront players with networks, only the latter actually includes monetary incentives. This hypothesis is tested by comparing the means of the decision scores of both in all decision parameters and showing that they are significantly different.

Do network incentives lead to self-regulation?

Before the relation between the incentives and self-regulation can be studied, first the notion of self-regulation must be translated into the context of the game. When is a network considered to be self-regulating? And how can self-regulation be quantified or at least partially ordered? Answering these questions with an exact metric is infeasible, hence two hypotheses are formulated that together approximate the original research question. The first is that due to the presence of incentives, the observed coordination level of every game is at least "Medium", which corresponds to a network-wide coordination of interactions. In games with a "Low" rating, coordination arises from conflict that need to be resolved between two or three players and is not considered coordination of the network. Secondly, in a self-regulated network, members are expected to act more in favour of the network, which is expressed in the game through a traffic or risk-averse play style. Thus, the hypothesis is that a strong positive correlation exists between the coordination level and the strategy scores for ttl G_t and risk-aversion G_r.

It must be noted that while the ultimate goal of the hypothesis is to prove a causal relationship between the use of monetary incentives and self-regulation ("leads to"), the limited setting of the serious game and its observations will at best reveal a correlation between the two. Still, the existence of such a correlation would be of great interest and, if found, future work can address the causality of this relationship.

Do relationships influence the effectiveness of incentives and, indirectly, self-regulation?

To confirm that decision making is influenced by social cohesion and not a predisposed preference of individual participants, first a comparison is performed between the a priori profile scores and the in-game strategy scores. The hypothesis is that there exists no correlation between all pairs Q_x and G_x, where x represents the profit, ttl and risk-aversion objectives. Assuming that this first hypothesis is confirmed, the inverse is studied for social cohesion. That is, it is hypothesised that social cohesion is strongly correlated to changes in strategy scores only and the change in preferences is statistically significant. Finally, it is to be expected that players that are more familiar with each other are more likely to coordinate their operations, expressed in a strong correlation between cohesion and coordination.

The road maintenance game

The hypothesis of the previous section are validated in a controlled simulated environment based upon the Maintenance Planning Problem of Scharpff *et al.* (2013). In this problem contractors must plan their maintenance work while they are accounted for their joint impact on traffic conditions. This paper simulates the application of a performance-based contract in this domain in the form of a serious game called the Road Maintenance Game. In this game, players take on the role of one of the five contractors known as "service providers" (SPs) that need to plan and execute maintenance work for the client, the "asset manager" (AM), on the road network visualised in Figure 2. As monetary incentives to incite self-regulation, the original network payments of Scharpff *et al.* (2013) are used that charges SPs 1 euro per additional hour of ttl caused. The design of the game entails a complex model of actions, rewards and rules, based upon figures of actual road maintenance projects from Brandt (2011), of which an elaborate description can be found in the on-line appendix (Scharpff *et al.* 2019). Here the core parts of the game are explained to prepare the reader with sufficient knowledge to understand the methodology and value the obtained results correctly.

Each player is responsible for an equivalent portfolio of 4 maintenance projects, for which they must select one of the four alternative methods to perform the work – "low cost", "low hindrance", "low risk" and "fast completion" – and plan them in time. The methods differ in their duration and direct impact on the
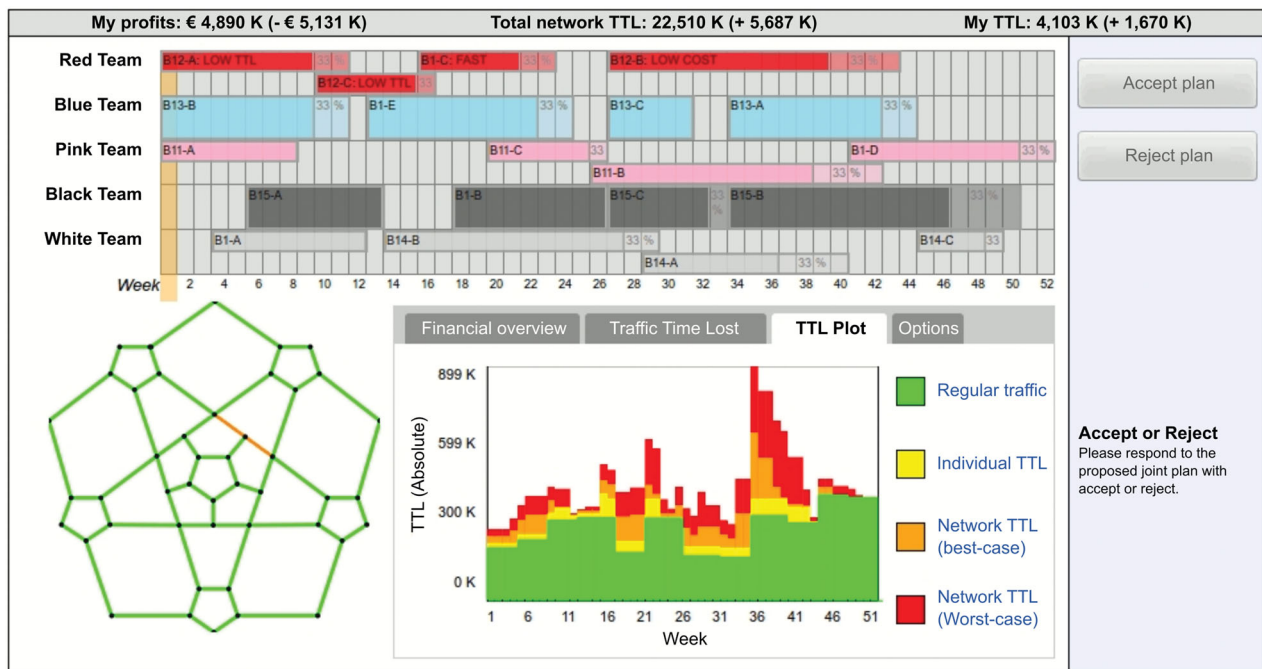
Figure 2. The user interface of a game in progress with the infrastructure (bottom left), the current joint plan as Gantt chart (top) and a graph showing its impact on ttl (bottom right). In the action bar on the right are the actions currently available to the player (Red Team).
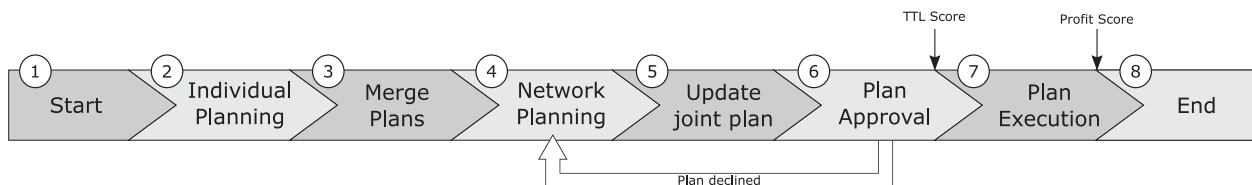

Figure 3. The various stages of the game and their succession.

profits and traffic hindrance, but also in their indirect impact in the form of risk of maintenance delays. All alternatives except for the "low risk" have a delay probability of 33% and varying delay durations. If delayed, the project may incur additional costs and ttl. The alternatives also vary in network impact: whereas the "low hindrance" alternative results in low ttl even when it is planned concurrently with other works, the "low cost" alternative will lead to a substantial amount of ttl if planned concurrently with others. Finally, players may decide to pass up on a project altogether and miss out on the contracted revenue for that task but perhaps avoid steep ttl penalties. During the game, the interface of Figure 2 functions as a decision-support system that computes and displays the impact of their decisions, incorporating the last-known decisions made by other players.

The game is run in a client-server fashion, where the game master (the AM) hosts a game server and all players (the SPs) connect to the game with Tablet PCs.

Every tablet corresponds to one of the players and shows the planning interface of which an example is shown in Figure 2. This interface visualises the planning decisions and their impact on profit and ttl in every stage of the game and presents the available actions to the players.

The gameplay follows the process flow of Figure 3:

1. At the start of the game players are introduced to the network, briefed and finally assigned one of the five maintenance portfolios for which they will be responsible during the rest of the game.
2. In the 'hedividual planning roundi they are given the task to develop a maintenance schedule according to their preferences. Their scheduling decisions can be summarised into two actions for every maintenance project: a) choose the preferred maintenance method and plan it in time or b) choose to not plan the method. Each player

Table 2. Outline of game session characteristics, from left to right the columns are: session identifier, company/institute, number of participants, number of questionnaires reviewed, participants skill category, and the observed coordination and social cohesion of participants (the categories are defined in Table 1).

	Company/institute Profile	#P	#Q	Category	Coordination	Cohesion
A	University, Computer Science	9	9	Students	Low	Unfamiliar
B	ICT-focussed R&D Company	10	9	Engineers	Low	Familiar
C	Utility provider, mainly power	15	3	Professionals	Low	Unfamiliar
D	Dutch national road authority	17	16	Trainees / Interns	High	Familiar
E	Dutch national road authority	8	5	Trainees / Interns	Medium	Familiar
F	AM Professionals Course	20	9	Professionals	Medium	Unfamiliar
G	AM and Health-care Consultants	16	9	Professionals	High	Familiar

individually submits their schedule to the AM when the schedule is satisfactory.

3. Once all players submitted their plans, the AM merges them into a single joint plan and sends that plan to all players thus informing them about the decisions made by others.

4. Now a 'owtwork planning roundi starts. With the newly received information about the plans of other players, every player is again requested to submit a maintenance schedule for their operations. This schedule can be the same as before, a slightly modified one to account for the other players or a completely new one. Once they are again satisfied with their schedule, they submit it to the AM and wait for the other players to do the same. Note that there are no real-time updates of the other playersm plans during this round, only when all plans are again submitted will this information be updated.

5. Once the (new) plans have again been received by the AM, the joint maintenance plan is updated and sent back to the players. This time an approval round is requested from all players.

6. In the approval round, every player either accepts or declines the joint plan. If any of the players declines, a new network planning round is started and the process is reset to the network planning round of step 4.

7. If all players accept the joint plan, the planning phase ends. From this point onward no more changes can be made to the joint maintenance schedule and the (expected) group ttl score is recorded. Now the execution phase starts, and the only action left in the game is the "realisation" of outcomes of the maintenance projects. In the execution phase, the plan execution is (gradually) simulated one week at a time until a maintenance operation starts that may delay. The player to which the task belongs then rolls a dice. If the dice lands on a green square, there will be no

delay in the execution of the task, whereas a red square means that the task is delayed (effectuated by the game master). This process is continued until all tasks have been fulfilled and a year has passed in the game. Then the game ends.

8. At the end of the game, the session winner is the player that has the highest profit after the execution of the joint maintenance plan.

Players can win the game in two ways. In every session, the player that has the highest profit after execution of the joint maintenance plan is declared the winner of the session. On the other hand, the group of players that achieves the lowest expected traffic time lost compared to all other sessions will win as a group (before execution). These two ways to win mimic the typical misalignment between the goals of the service providers (maximum profit) and that of the contractor (minimal nuisance) that is seen in practice. To emphasise this misalignment and provoke competition, there is only a small price for the winning player in a single session (€2.50 scratch ticket) but a more valuable one for the players in the winning session (€10 vouchers for all participants of the session).

Gaming results

This section presents visual and tabular summaries of the data gathered through the combination of the questionnaire and the Road Maintenance Game. The full data set can be found in the on-line appendix (Scharpff et al. 2019) and the GitLab repository[2] of the first author. References to the data set are included.

For the sessions both public institutions and companies were contacted that focus on asset-management related activities. Getting together a large enough group of people able to participate in a gaming session of approximately 3 h proved to be challenging, especially when dealing with practitioners from the industry. Nonetheless, 7 sessions with a total of 95 participants were hosted in groups of varying
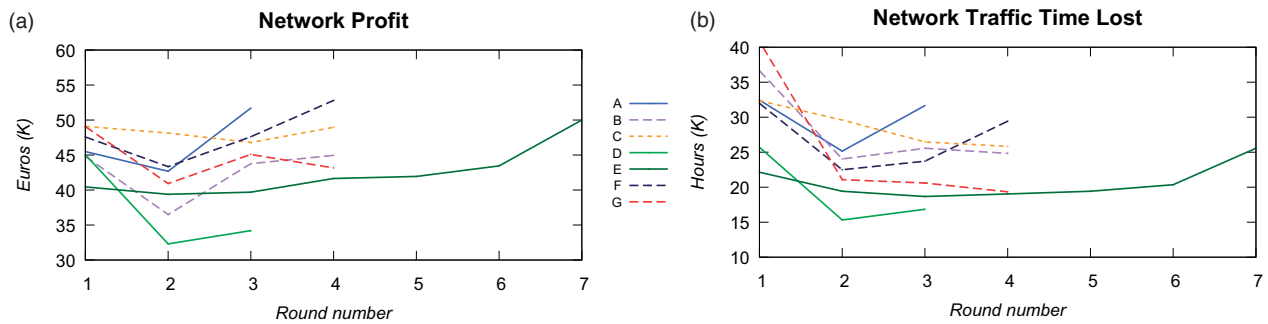
Figure 4. Quality of the joint plan at the end of each round in terms of the goals profit and ttl: (a) shows the expected profit in thousands of euros, (b) the expected ttl in thousands of hours. Note that the ranges on the vertical axes do not align, but the sizes of the ranges do. The line styles associated with each session are shown in the legend in the middle.
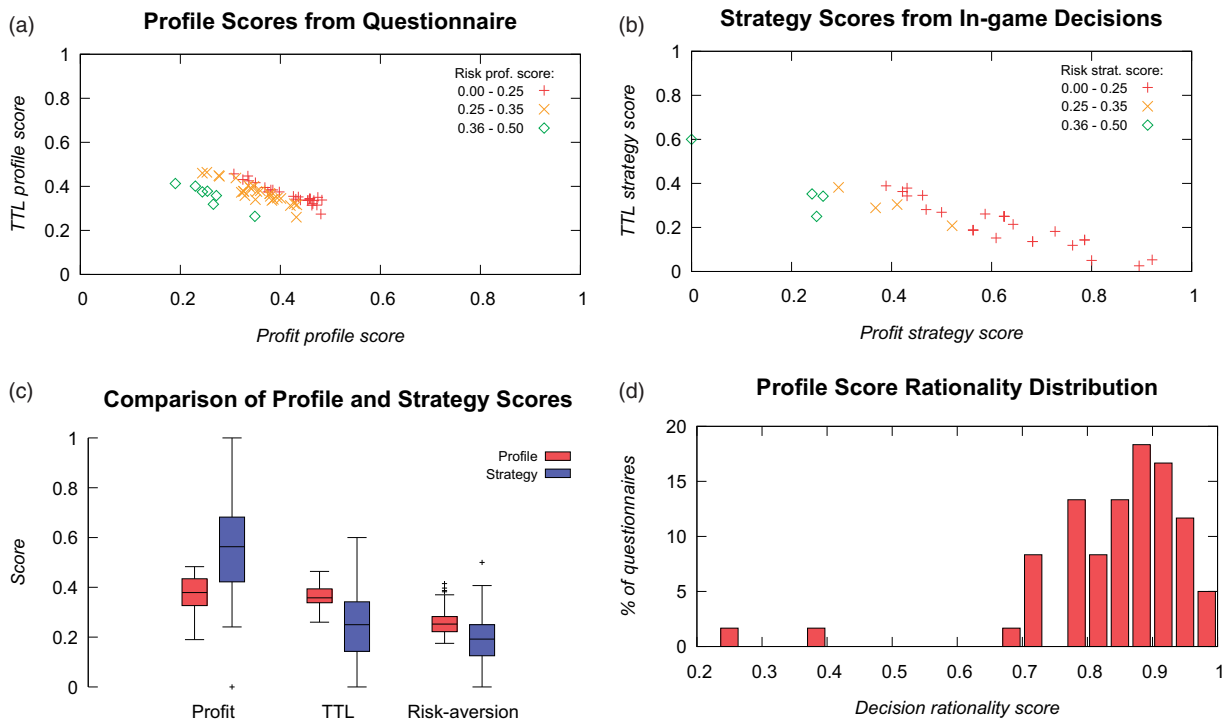

Figure 5. Visualisation of (a) the profile scores determined from the questionnaire responses, (b) the strategy scores computed from in-game action decisions, (c) a comparison of their distributions and (d) the computed rationality of questionnaire responses.

composition, skill and social coherence. Furthermore, 60 questionnaire responses were received from the participants. Table 2 gives a descriptive overview of all sessions. The overall results of all sessions in terms of the game goals, i.e. maximising profit and minimising ttl, are summarised in the graphs of Figure 4. The figures show the expected profit and ttl of the group as a whole, as recorded at the end of each round (Section 3.2 of the appendix).

Outcomes and decision-preference scores are computed over $N-1$ rounds, such that N is the number of rounds played in a session. This is to correct for the "last round" effect due to the design of the game that the player with the highest expected profit *at the end of the game* wins the session. This caused players to

radically change their strategy in the last round to a profit-driven one in an effort to win the prize. With the exception of session *G*, all games suffered from this effect. This can also be observed from the graphs in Figure 4: all lines show a substantial growth in profits in the last round, often paired with an increase in ttl.

The decision preferences of participants are visualised through the graphs in Figure 5, where Figure 5(a) plots the profile scores and Figure 5(b) the strategy scores. Both figures show the profit and ttl preferences on the x and y axis respectively and use different point styles to categorise the risk aversion scores associated with each point. For example, points visualised by a diamond shape correspond to a preference score

with a preference score for risk aversion between 0.36 and 0.5. Risk aversion scores greater than 0.5 are not found in either the profile score or the strategy score. Figure 5(c) visualises the distribution of profile scores and strategy scores using a box plot. This picture illustrates the range in which preferences are typically expressed. The means are illustrated by the thick lines in the vertical middle of the boxes, that represent the first and third quartile of the data set. The whiskers visualise the most distant point on both ends that are at most 1.5 away from the inter-quartile distance. Finally, the histogram of Figure 5(d) shows the distribution of rationality as computed from the questionnaire responses. All visualisations use the scores from Table 3.2 and Section 3.3 of the appendix.

Based upon the same data sets, a correlation analysis is performed on the relationship between profile scores and strategy scores. Correlation coefficients are computed using the Pearson method (Taylor 1990, Zou *et al.* 2003) and correlation confidence is computed through two-tailed sample t-tests with confidence levels 95% and 99%. Interpretation follows the model for social sciences (Cohen 1992) and conform the labelling by Taylor (1990) defined as weak for coefficient values in the range [0, 0.35], moderate for [0.36, 0.67] and strong for [0.68, 1.0]. The summary of this analysis is shown in Table 3.

To gain insight into the interactions between the parameters in the environment and their relation to the outcomes of the game a comprehensive correlation analysis is performed, of which the summary is presented in Table 4. Correlations shown in bold correspond to a confidence level of at least 99%, italic to 95% and the others are shown in grey. For the game outcomes, the expected profit P and ttl T is listed along with the associated performance ratio ϕ and the worst-case profit P_w and ttl T_w with their performance ratio ϕ_w. The abbreviations CD and CH refer to respectively the coordination level and the cohesion level of games. The figures underlying this table can be found in Table 3.1 (coordination and cohesion level), Table 3.2 (profile scores), Section 3.3 (strategy scores) and Table 3.3 (outcomes) of the appendix.

Table 5 takes a closer look into the variables coordination level and cohesion level. This table lists the averages per category for the expected outcomes and

Table 3. The correlation strengths and associated confidence levels (grey) for every pair of profile and strategy scores.

	Q_p		Q_t		Q_r	
G_p	−0.10	57%	−0.10	54%	0.22	91%
G_t	0.14	73%	0.12	63%	−0.29	98%
G_r	0.00	2%	0.15	75%	−0.13	68%

Table 4. Summary of the correlation analysis between the variables and outcome.

Variable	Expected outcomes				Worst-case outcomes					
	P	T	u	ϕ	P_w	T_w	u_w	ϕ_w	CD	CH
Q_p	−0.15	*0.31*	**−0.47**	**−0.34**	−0.13	−0.19	0.13	0.07	**−0.51**	−0.18
Q_t	−0.01	−0.13	0.13	0.14	−0.11	−0.15	0.11	0.07	0.10	0.02
Q_r	0.21	*−0.30*	**0.52**	**0.34**	*0.27*	**0.38**	*−0.27*	*−0.15*	**0.60**	0.22
G_p	**0.87**	**0.64**	0.09	*−0.41*	**0.80**	**0.68**	−0.31	*−0.43*	−0.37	**−0.69**
G_t	**−0.82**	**−0.53**	−0.18	0.28	**−0.81**	**−0.55**	0.13	0.24	0.26	**0.68**
G_r	**−0.78**	**−0.67**	0.05	**0.50**	**−0.63**	**−0.72**	**0.48**	**0.59**	**0.58**	**0.58**
Coordination	−0.23	*−0.45*	*0.37*	**0.59**	−0.12	**−0.51**	**0.58**	**0.68**	–	**0.57**
Cohesion	**−0.57**	*−0.42*	−0.05	0.33	**−0.50**	**−0.47**	0.25	*0.38*	**0.57**	–

All correlation coefficients with a statistical confidence level greater than 99% are shown in bold, greater than 95% in italic and all correlations with lower likelihoods of being correct predictors are depicted in grey.

Table 5. Average values for the outcomes and scores per coordination and cohesion category.

	Expected outcomes				Worst-case outcomes				Profile scores			Strategy scores		
	P	T	u	ϕ	P_w	T_w	u_w	ϕ_w	Q_p	Q_t	Q_r	G_p	G_t	G_r
Coordination														
Low	45.7	28.6	17.1	1.62	33.0	35.2	−2.1	0.95	0.43	0.36	0.21	0.60	0.23	0.18
Medium	43.3	23.6	19.6	1.87	31.9	28.1	3.8	1.16	0.35	0.37	0.28	0.53	0.23	0.24
High	43.2	22.4	20.8	2.00	31.8	26.1	5.7	1.27	0.34	0.37	0.29	0.50	0.27	0.24
Correlation	−0.23	*−0.45*	*0.37*	**0.59**	−0.12	**−0.51**	**0.58**	**0.68**	**−0.51**	0.10	**0.60**	*−0.37*	0.26	**0.58**
Cohesion														
Unfamiliar	47.7	28.3	19.4	1.70	35.0	34.7	0.3	1.02	0.39	0.37	0.24	0.65	0.18	0.17
Familiar	42.0	23.1	18.9	1.90	30.7	27.2	3.4	1.18	0.36	0.37	0.27	0.48	0.28	0.24
Correlation	**−0.57**	*−0.42*	−0.05	0.33	**−0.50**	**−0.47**	0.25	*0.38*	−0.18	0.02	0.22	**−0.69**	**0.68**	**0.58**

Correlation coefficients with a confidence level of at least 99% are shown bold, 95% in italic and the rest in grey.

associated performance ratio, the worst-case outcomes, profile scores and strategy scores of either variable. The same data set as for Table 4 was used in a similar fashion to construct this table. Note that the correlation coefficient included in this table are the same as in Table 4 but restated to provide a more complete overview.

Findings

Using the summaries of the previous section, the research questions are again considered. From the measurements and observations, the hypotheses formulated for the research questions are tested. The same structure as before is used in presenting the findings relevant to each question.

Validity of the model

To assert that human decision makers are at least boundedly rational, the hypothesis is that the mean of rationality scores is at least 0.8. In other words, human decision makers are capable in finding at least 80% optimal solutions on average. This assumption is validated by performing a one-tailed, one-sample t-test where the mean of rationality scores μ_θ is compared against a mean of 0.8 using null-hypothesis $H_0 : \mu_\theta > 0.8$ and a confidence level of 95%. The resulting probability value is 0.010 and since this is lower than the desired confidence value of 0.05, it is concluded that the null-hypothesis is valid. Furthermore, there are two outliers at 0.255 and 0.385 that likely correspond to misunderstanding the questionnaire, as the rationality of the next lowest is close to 0.7. When these outliers are removed, the mean of rationalities becomes greater than 0.84 with the same confidence level.

The correctness of the game design is illustrated by the correlations between the strategy scores and the outcomes listed in Table 4. Starting with the profit strategy score G_p, the table shows a strong positive correlation between this score and the total expected profits P and worst-case profits P_w of the game outcome as the absolute value of the coefficient is greater than 0.67. Profit strategy scores show a moderate to strong positive correlation to the ttl in the resulting outcomes. Note that this positive correlation is conform the intended design as higher ttl values mean more network hindrance. For ttl, a similar analysis shows that it is strongly negatively correlated to profits, and moderately negatively correlated to ttl. Risk aversion is also strongly negative correlated with

both and in addition shows a moderately positive correlation with the performance ration, e.g. the ratio of profit versus ttl, thus the risk-averse actions are likely to lead to performance increases. In conclusion, although the original hypothesis of strong correlations is not proven, the correlation coefficients are all in the range of moderate to strong correlation and hence the hypothesis is highly plausible.

It remains to show that coordination is beneficial to the outcomes. Table 5 illustrates the correlations of interests. These figures do not provide sufficient evidence for the hypothesis of a strong correlation, nonetheless they do reveal a moderate correlation between the coordination level and the utilities. This suggest that again it is at least plausible that coordination is beneficial to the players, also supported by the observed average utility per category. Moreover, the performance ratio shows a moderate to strong correlation to the coordination level, indicating that more coordination is likely to improve the trade-off between profits and ttl. Finally, coordination seems to substantially improve the worst-case scenarios in the game to the benefit of all players.

Influence of monetary incentives on decision making

The effectiveness of the monetary incentives is measured through the hypothesis that a significant difference can be observed between the decision making with and without monetary incentives. The decision-making preferences in the former situation are given by the strategy scores from the game, the latter through the profile scores computed based on the questionnaire. Indeed, Figure 5(a) to Figure 5(c) appear to visualise a difference between the two categories. In Figure 5(a), the profile scores seem relatively balanced with a slight preference towards profit and ttl over risk aversion. The strategy scores of Figure 5(b) seemingly indicate that when monetary incentives are used, a profit-focussed play style is preferred. This assessment is supported by the box plot in Figure 5(c) that visualises the distributions of both score types. A two-tailed, paired t-test between the means of all Q_x and G_x is performed with null-hypotheses $H_0 : \mu_{Q_x} = \mu_{G_x}$ and a confidence level of 95% to confirm the statistical significance of the difference. The resulting probability values are 0.009, 0.001 and 0.313 for respectively the profit, ttl and risk-aversion scores. Hence, with high statistical likelihood, the changes in preference for profit and ttl can be ascribed to the presence of monetary incentives, but risk-aversion seems unaffected.

Self-regulation of the network

The results indicate that a change of behaviour can be observed in the presence of incentives. Hence it is interesting to know whether that change corresponds to the manifestation self-regulation. The first hypothesis in this regard is that the presence monetary incentives always leads to a coordination level of "Medium" or "High". A quick glance at Table 2 is sufficient to invalidate this hypothesis. In three games coordination was limited to only conflict-driven negotiations. The second hypothesis states that monetary incentives lead to a change in decision-making strategy towards a ttl-driven or risk-averse play style in the game. Table 5 however shows only a moderate correlation between coordination and risk aversion, and no such a relationship is established between coordination and ttl strategy scores. Hence the incentives are unlikely to incite self-regulation.

An interesting additional find from the same table is that whereas Table 3 illustrates that a correlation between profile scores and game scores is highly unlikely, the coordination level is moderately correlated to both the profile and strategy scores for risk. This does not reveal a relationship between risk aversion and coordination, but it does suggest that network members with a predisposition against risks are more likely to coordinate their operations, with or without the presence of incentives.

Role of social cohesion

As incentives do not consistently incite self-regulation, the influence of relationships on coordination of even greater interest. First the hypothesis is tested that preference changes are due to the presence of familiar network members and not from player profiles. The absence of any relationship between the profile scores and social cohesion can be concluded from Table 5 with a high likelihood. All profile parameters are at most weakly correlated with low likelihood (respectively 83%, 10% and 91% for profit, ttl and risk aversion). The strategy scores, on the other hand, appear to be related to cohesion. With risk aversion being at least moderately correlated, a strong correlation is revealed for profit and ttl strategy scores. A one-tailed paired t-test is performed to analyse the statistical significance of this change in strategy means. The use of a one-tailed test follows from the strong directions of the coefficients. As null-hypotheses $H_0 : \mu_{G_{x,U}} > \mu_{G_{x,F}}$ are used, such that U and F correspond to the categories "Unfamiliar" and "Familiar" and x to the objectives. The tests yield probability values 0.020, 0.028 and 0.035 for respectively the profit, ttl and risk-aversion strategy scores with a confidence level 95% and hence the behavioural change to a collaborative play style is statistically significant. With respect to the correlation between cohesion and coordination, a similar conclusion can be drawn as twice before. Table 4 provides evidence of a moderate correlation but the original hypothesis of a strong correlation is not satisfied, although a relationship is very plausible. Summarising these findings, social cohesion is a probable moderator for the effectiveness of monetary incentives.

Discussion

Here the findings of the experiments are discussed and their relevance to the broader context is addressed. First, the effectiveness of monetary incentives to influence decision making is considered and in particular how behaviour is changed and how this relates to current literature. Thereafter the role of social cohesion is investigated in a similar fashion.

The effectiveness of incentives to influence decision making was already highlighted by other scholars such as Bresnen and Marshall (2000), Bower et al. (2002) and Rose and Manley (2011). This study contributes an empirical confirmation that indeed incentives changes behaviour of participants within the controlled experimental setting and provides an insight into *how behaviour is influenced* (Figures (a) to (c)). The findings show that the payment mechanism used in this game led to a competitive play style focussed on profits, whereas the questionnaire responses of the same group of participants were more balanced in their decision preferences. Even more so, all but a few participants motivated their responses in the questionnaire among the lines of "this seems to optimally balance profits and ttl" when asked why they chose a specific alternative. In the game, however, profit was observed to be the key driver of almost every player. Nonetheless, the profit-driven behaviour did not result in the emergence of coordination, even while players could improve their utility by coordinating their actions. Further investigation into the existence of any correlation between coordination and agent decision making reveals moreover that no substantial differences are found in the strategic preferences for different levels of coordination, except for a stronger risk aversion (Table 5). The latter effect is explained by an increased focus on robust planning so that coordination of activities is guaranteed. The experiments here therefore contribute to an improved understanding of the influence of incentives on decision making, but also lead to wondering why coordination does not emerge, even if players can benefit from doing so.

One explanation that is offered by literature is the inability of players to fully comprehend the complexity of the domain and thus fail to maximise their gains, as observed by for instance Eriksson (2010) and Shadid (2018). Although the participants performed near optimal in their questionnaire responses, the complex dynamics of a coordinating a 5-player network under pressure of time via a new interface may justify this explanation. Along the same lines is the relative novelty of the role the players assume which requires an additional set of skills and capabilities (Straub 2010, Klijn et al. 2010b, Hartmann and Hietbrink 2013) that the participants may not possess. In either case, a mitigating measure is to employ computer-aided decision support techniques that help to maximise the value of planning and suggest coordination to network members when this is beneficial, leading even the most isolated or selfish network members to coordinate decisions. Positive examples in similar decision-support scenarios can be found in the works by e.g. Cheung et al. (2005), Le Bars and Le Grusse (2008) and Douma et al. (2012). Of additional value is the use of collaboration tools and knowledge sharing, as proposed by Eriksson (2010). Whether these extra measures result in more coordination could be established through additional gaming experiments and would be a valuable empirical contribution by future work.

A more plausible explanation offered by e.g. Bresnen and Marshall (2000) and Fehr and Falk (2002) is that the monetary incentives lead to an effect that is opposite to their intended design. The payment mechanism used in the game penalises the players per additional hour of traffic time lost. Although the mechanism was intended to stimulate reduction of hindrance due to maintenance and encourage efficient and innovative operations, it can be experienced as a painshare mechanism that only penalises bad behaviour. This type of mechanism has been found much less effective to motivate contractors. For instance, Choi et al. (2011) conclude from a large survey of completed projects that agreements that incorporate both positive and negative incentives successfully improved performance while agreements with only penalties led to performance worse than conventional contracts. Through a serious gaming study not unlike the one performed here, Altamirano and de Jong (2009) demonstrated that high penalties seem to create incentives for collusive behaviour, while a combination of moderate penalties with significant bonuses creates a positive atmosphere of trust. Similarly, Hosseinian and Carmichael (2013) remark that "sharing gain/pain provides a strong motivational factor for all

parties to work together, rather than in a confrontational or adversarial fashion, with the desired result of producing a successful project". Particularly the "adversarial fashion" seems to fit the observations made here. Instead of motivating players to collaborate, the incentives appeared to stimulate a competitive attitude with a decrease in ttl and risk-averse preferences in favour of a profit-driven play style. Not only is competition within the network regarded detrimental to its performance (Conrad et al. 2003, Shadid 2018), it is counterproductive to an open and cooperative network environment in which decision coordination and co-creation is stimulated (Hosseinian and Carmichael 2013, Hartmann et al. 2014).

In contrast, the experimental results regarding social cohesion present a new empirical confirmation of the positive impact of the social dimension on network performance. A contribution of this work is that while networks with familiar members demonstrate a more collaborative attitude in their decision making, as per expectation, it is newly shown that no such inclination was observed in questionnaire responses of the same participants. Another new finding is that while the "socially-cohesive" networks were confronted with the same monetary incentives as the unfamiliar networks, they did not show a similar competitiveness. As a corollary, the presence of familiar network members is seemingly influencing their decision making towards a more collaborative style and supports the incentive mechanism in achieving its intended results. This once more stresses the importance of building a socially cohesive network, as many authors have before (Bresnen and Marshall 2000, Agranoff and McGuire 2001, Klijn et al. 2010a, Volker et al. 2014), but contributes also the learning that relationships within the network may be a necessary condition for (inadequately designed) monetary incentive schemes, or are at least beneficial to the goal of maximising network performance.

Limitations and future work

Ultimately, the failure of monetary incentives to incite coordination in every gaming session may be interpreted as evidence against the use of incentives to achieve self-regulation, but a closer look at the results suggests otherwise. The experiments in a controlled setting reveal a definite potential for monetary incentives but their potential is determined by the type of incentives that are used, the relationships within the network or a combination of both. From the gaming sessions it is yet impossible to conclude the exact role

of either element. As a consequence, further study is strongly recommended to isolate either the design of the mechanism or social cohesion as the main contributor to self-regulation, or conclude that both are prerequisite to network self-regulation. Regardless, the experiments performed provide new insights and recommendations for both the research on and the application of incentives in networks. With the next step of bringing self-regulation into practice through monetary incentives, and evaluating their effect in real-world scenarios, three main directions for current and future work are identified.

First of all, adequate design of the incentive scheme is vital to the success of performance-based tenders and care should be taken in the type of incentives that are implemented, especially when good relationships cannot be guaranteed. The use of painshare without the gainshare led to competition and selfish optimisation within the network. This learning should be considered in game-theoretically engineered incentive schemes such as that by Gupta et al. (2011), Scharpff et al. (2013), Volker et al. (2012) and Hong et al. (2016). Although theoretically such mechanisms should result in optimal coordination of the network, in practice they can be counterproductive to self-regulation if they rely on incompatible payment mechanisms such as painshare-only schemes. The effect of incentive design, and in particular the type of mechanism employed, needs further investigation to establish the exact mediators that realise the desired changes in decision making towards self-regulation.

Secondly, the findings here again underline the paramount importance of the social dimension in network-based approaches and emphasise that relationships between network members are to be fostered. Even though the precise effect of social cohesion is not fully determined, monetary incentives to stimulate self-regulation seem most effective in cooperative networks. Its promise is hence the greatest in managing the interactions in collaborative networks such as alliances, early contractor involvements or integrated project delivery. In such networks cooperative behaviour between the parties is established and measures are taken to get all parties to work to the same goal Carmichael 2000, Hosseinian and Carmichael 2013. Moreover, alliances are becoming the preferred delivery for large and complex projects (Hauck et al. 2004, Love et al. 2011) and hence the contribution of self-regulation as a method to optimise network performance is very relevant for future work.

Thirdly, to bring self-regulation into realistic tenders, concerns will have to be addressed with respect to accountability and the possibility of failure to stimulate desired performance. While the application of performance-based contracting on the network level is certainly promising, the lack of a principal in the network may lead to gaps in the responsibility of the network (Agranoff and McGuire 2001). Self-regulation is inherently in tension with the accountability of the network and requires a solid legal basis before practical use can be realised. Furthermore, as noted by Turrini et al. (2010) and demonstrated by the experiments, decentralised tools and lack of coordination by a central institution or agency can certainly worsen the performance and output of a network and may in the worst case lead to a fall-back to regulatory approaches (Snippert et al. 2015). More research is needed into mechanisms and their enforcement to prevent such fall-backs in practice. One typical approach to reduce the probability of failure and increase the effectiveness of incentives is to perform regular monitoring (Javed et al. 2014, Gao and Liu 2019), or at least increase the perception thereof (Nalbantian and Schotter 1997), but this is typically paired with substantial costs to the client.

Finally, care must be taken in the interpretation of the results from the experiments and their translation to the real world. Although players often show behaviour similar as they would in comparable real-world situations, the game will always remain a simplified model of the world and hence behaviour cannot be directly extrapolated to realistic scenarios. The absence of real-world consequence may also cause players to take more risks than they normally would. The gaming setting however helps to explore potential changes in decision-making preferences and patterns in behaviour under controlled conditions that may guide further experimentation and research. On that note, the serious game itself is an additional contribution of this work and available on-line for future research. It can be used for instance to play the same game with different incentive mechanisms, or the absence of incentives, to further investigate their impact, but also to explore the potential related approaches such as the Dynamic Contracting framework of Volker et al. (2012) or the gainshare/painshare mechanisms of Hosseinian and Carmichael (2013). Alternatively, it can be employed as a tool to get practitioners acquainted to the concept of self-regulation, alike Altamirano et al. (2008), Harteveld et al. (2010) and Douma et al. (2012), or to perform many experiments using automated agents to play the game (Wenzler and Chartier 1999)

Conclusion

While many authors such as Bresnen and Marshall (2000), Smith and Grinker (2004) and Volker *et al.* (2014) advocate the benefits of innovative performance-based contracts in the context of construction works, the effectiveness of monetary incentives to achieve self-regulation within an infrastructure network is uncertain. On this note, the study contributes on what can be empirically expected if self-regulation is adopted by simulating the application thereof in a controlled serious game experiment. The results show that incentives are effective in changing decision preferences but may lead to counterproductive effects. Instead of inciting better performance, the findings demonstrate that players are driven to a competitive attitude, detrimental to collaboration. Furthermore, it was observed that networks with strong social cohesion can overcome this competition and will self-regulate their performance as intended by the incentive design. These results confirm previous findings on the importance of relationships within the network and additionally signify the role of social cohesion on the effectiveness of incentive mechanisms, suggesting that strong social cohesion is prerequisite to performance-based monetary incentive mechanisms.

The recommendation ventured from this study is that if self-regulation is to be successfully implemented in contracts, it should be based on positive incentives structures that incite better performance in a cooperative atmosphere of co-creation and trust. Consequentially, self-regulation is deemed most effective in collaborative team settings such as alliances and early contractor involvements or integrated project delivery. Future work is needed to clarify the impact of the incentive design on self-regulation, compare the findings here to the non-incentivised setting and perform repeated studies to strengthen the conclusion drawn in this study on road maintenance. The serious game setup of this study can be reused in future settings and is available on-line to be used in related studies.

Disclosure statement

The data that support the findings of this study are openly available from the first author's GitLab repository at https://gitlab.com/jscharpff/maintenance-planning-game. The source code for the Maintenance Planning Game and the rationality computation can be found within the same repository.

Notes

1. Traffic time lost expresses the hours of travelling time lost due to the congestion of the road network. See for instance the reports by Brandt (2011) or the papers by Altamirano and de Jong (2009) and Scharpff et al. (2013)
2. https://gitlab.com/jscharpff/maintenance-planning-game

ORCID

Daan Schraven http://orcid.org/0000-0003-0647-1172
Leentje Volker http://orcid.org/0000-0003-2766-3763
Mathijs M. de Weerdt https://orcid.org/0000-0002-0470-6241

References

Agranoff, R., and McGuire, M., 2001. Big questions in public network management research. *Journal of public administration research and theory*, 11 (3), 295–326.

Altamirano, M.A., and de Jong, W.M., 2009. Opportunistic behavior in road maintenance markets. *Transportation research record: journal of the transportation research board*, 2108 (1), 13–22.

Altamirano, M., Herder, P., and De Jong, M. (2008). Road roles using gaming simulation as decision technique for future asset management practices. *In: Proc. of the International Conference on Systems, Man and Cybernetics*, Singapore, pp. 2297–2302.

Axelrod, R., 2003. Advancing the art of simulation in the social sciences. *Japanese journal for management information system, special issue on agent-based modeling*, 12, 1–19.

Bower, D., *et al.*, 2002. Incentive mechanisms for project success. *Journal of management in engineering*, 18 (1), 37–43.

Brandt, M. H. (2011). Kwartaalmonitor bereikbaarheidsontwikkeling hoofdwegennet 1e kwartaal 2011. Ministerie van Infrastructuur en Milieu Rijkswaterstaat, Dienst Verkeer en Scheepvaart, Technical report. Available from: http://publicaties.minienm.nl/documenten/kwartaalmonitor-bereikbaarheidsontwikkeling-hoofdwegennet-1e-k-2 [Accessed 28 April 2019] (Dutch only).

Bresnen, M., and Marshall, N., 1998. Partnering strategies and organizational cultures in the construction industry. In: Proceedings, ARCOM 14th Annual Conference, University of Reading, 2, 465–476.

Bresnen, M., and Marshall, N., 2000. Motivation, commitment and the use of incentives in partnerships and alliances. *Construction management and economics*, 18 (5), 587–598.

Calderón, A., and Ruiz, M., 2015. A systematic literature review on serious games evaluation: An application to software project management. *Computers & education*, 87, 396–422.

Carmichael, D. G., 2000. *Contracts and international project management*. Boca Raton, FL: CRC Press.

Cheung, W., Leung, L.C., and Tam, P.C., 2005. An intelligent decision support system for service network planning. *Decision support systems*, 39 (3), 415–428.

Choi, K., *et al.*, 2011. Schedule effectiveness of alternative contracting strategies for transportation infrastructure improvement projects. *Journal of construction engineering and management*, 138 (3), 323–330.

Cohen, J., 1992. Statistical power analysis. *Current directions in psychological science*, 1 (3), 98–101.

Conrad, D.A., *et al.*, 2003. Community care networks: Linking vision to outcomes for community health improvement. *Medical care research and review: MCRR*, 60 (4 Suppl), 95S–129S.

Dash, R.K., Jennings, N.R., and Parkes, D.C., 2003. Computational-mechanism design: A call to arms. *IEEE intelligent systems*, 18 (6), 40–47.

Demirel, H.Ç., *et al.*, 2017. Flexibility in ppp contracts–dealing with potential change in the pre-contract phase of a construction project. *Construction management and economics*, 35 (4), 196–206.

Dewulf, G., and Kadefors, A., 2012. Collaboration in public construction-contractual incentives, partnering schemes and trust. *Engineering project organization journal*, 2 (4), 240–250.

Douma, A.M., van Hillegersberg, J., and Schuur, P.C., 2012. Design and evaluation of a simulation game to introduce a multi-agent system for barge handling in a seaport. *Decision support systems*, 53 (3), 465–472.

Eisenhardt, K.M., 1985. Control: Organizational and economic approaches. *Management science*, 31 (2), 134–149.

Eriksson, P.E., 2010. Partnering: what is it, when should it be used, and how should it be implemented? *Construction management and economics*, 28 (9), 905–917.

Fehr, E., and Falk, A., 2002. Psychological foundations of incentives. *European economic review*, 46 (4-5), 687–724.

Gao, R., and Liu, J., 2019. Selection of government supervision mode of PPP projects during the operation stage. *Construction management and economics*, 37(6), 1–20.

Gigerenzer, G., and Selten, R., 2002. *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gupta, D., Vedantam, A., and Azadivar, J., 2011. *Optimal contract mechanism design for performance-based contracts*. Minnesota Department of Transportation Research Services Section, Technical report, St. Paul, MN.

Guyot, P., and Honiden, S., 2006. Agent-based participatory simulations: merging multi-agent systems and role-playing games. *Journal of artificial societies and social simulation*, 9 (4).

Harteveld, C., *et al.*, 2010. Balancing play, meaning and reality: the design philosophy of levee patroller. *Simulation & gaming*, 41 (3), 316–340.

Hartmann, A., and Hietbrink, M., 2013. An exploratory study on the relationship between stakeholder expectations, experiences and satisfaction in road maintenance. *Construction management and economics*, 31 (4), 345–358.

Hartmann, A., *et al.*, 2014. Procuring complex performance: the transition process in public infrastructure. *International journal of operations & production management*, 34 (2), 174–194.

Hauck, A.J., *et al.*, 2004. Project alliancing at national museum of australia: collaborative process. *Journal of construction engineering and management*, 130 (1), 143–152.

Hong, S., Wernz, C., and Stillinger, J.D., 2016. Optimizing maintenance service contracts through mechanism design theory. *Applied mathematical modelling*, 40 (21–22), 8849–8861.

Hosseinian, S.M., and Carmichael, D.G., 2013. Optimal gain-share/painshare in alliance projects. *Journal of the operational research society*, 64 (8), 1269–1278.

Hughes, W., and Kabiri, S., 2013. *Performance-based contracting in the construction sector*. University of Reading, Technical report, United Kingdom.

Javed, A.A., Lam, P.T., and Chan, A.P., 2014. Change negotiation in public-private partnership projects through output specifications: an experimental approach based on game theory. *Construction management and economics*, 32 (4), 323–348.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *The American psychologist*, 58 (9), 697

Klijn, E.-H., Edelenbos, J., and Steijn, B., 2010a. Trust in governance networks: its impacts. *Administration & society*, 42 (2), 193–221.

Klijn, E.-H., Steijn, B., and Edelenbos, J., 2010b. The impact of network management on outcomes in governance networks. *Public administration*, 88 (4), 1063–1082.

Kuitert, L., Volker, L., and Hermans, M.H., 2019. Taking on a wider view: public value interests of construction clients in a changing construction industry. *Construction management and economics*, 37 (5), 257–277.

Lam, T., and Gale, K., 2014. Highway maintenance: impact of framework agreements upon project financial performance. *Construction management and economics*, 32 (5), 460–472.

Le Bars, M., and Le Grusse, P., 2008. Use of a decision support system and a simulation game to help collective decision-making in water management. *Computers and electronics in agriculture*, 62 (2), 182–189.

Love, P.E., *et al.*, 2011. Risk/reward compensation model for civil engineering infrastructure alliance projects. *Journal of construction engineering and management*, 137 (2), 127–136.

Meijer, S.A., *et al.*, 2012. Gaming rail cargo management: exploring and validating alternative modes of organization. *Simulation & gaming*, 43 (1), 85–101.

Nalbantian, H.R., and Schotter, A., 1997. Productivity under group incentives: an experimental study. *The American economic review*, 87(3), 314–341.

Nisan, N., *et al.*, 2007. *Algorithmic game theory*. Cambridge, United Kingdom: Cambridge University Press.

Parkhe, A., 1993. Strategic alliance structuring: a game theoretic and transaction cost examination of interfirm cooperation. *Academy of management journal*, 36 (4), 794–829.

Potemans, A., Volker, L., and Hermans, M. (2018). From agents to stewards? Experiences from a dutch infrastructure case study. *In: 34th Annual ARCOM Conference*. ARCOM, Association of Researchers in Construction Management. Belfast, Ireland.

Rose, T., and Manley, K., 2011. Motivation toward financial incentive goals on construction projects. *Journal of business research*, 64 (7), 765–773.

Roszkowska, E. (2013). Rank ordering criteria weighting methods–a comparative overview. *Optimum. Studia Ekonomiczne*, 5 (65): 14–33.

Roth, A.E., 2002. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70 (4), 1341–1378.

Rousis, D. (2011). *A Pareto frontier intersection-based approach for efficient multiobjective optimization of competing concept alternatives*. PhD thesis. Georgia Institute of Technology.

Scharpff, J., et al. (2019). *The road maintenance planning game – game design and first results*. Delft University of Technology, Technical report. Delft, The Netherlands.

Scharpff, J., et al. (2013). Planning under uncertainty for coordinating infrastructural maintenance. In: *Proc. of the International Conference on Automated Planning and Scheduling*. Rome, Italy.

Schraven, D., Hartmann, A., and Dewulf, G., 2011. Effectiveness of infrastructure asset management: challenges for public agencies. *Built environment project and asset management*, 1 (1), 61–74.

Shadid, W.K., 2018. A framework for managing organizations in complex environments. *Construction management and economics*, 36 (4), 182–202.

Smith, D. C., and Grinker, W. J., 2004. *The promise and pitfalls of performance-based contracting*. New York City, NY: Seedco.

Snippert, T., et al., 2015. Barriers to realizing a stewardship relation between client and vendor: the best value approach. *Construction management and economics*, 33 (7), 569–586.

Straub, A., 2010. Competences of maintenance service suppliers servicing end-customers. *Construction management and economics*, 28 (11), 1187–1195.

Taylor, R., 1990. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6 (1), 35–39.

Tesfatsion, L., and Judd, K. L., 2006. *Handbook of computational economics: agent-based computational economics*. Amsterdam, The Netherlands: Elsevier.

Triantaphyllou, E., 2013. *Multi-criteria decision making methods: a comparative study*, volume 44. Berlin, Germany: Springer Science & Business Media.

Turrini, A., et al., 2010. Networking literature about determinants of network effectiveness. *Public administration*, 88 (2), 528–550.

Van Bueren, E.M., Klijn, E.-H., and Koppenjan, J.F., 2003. Dealing with wicked problems in networks: Analyzing an environmental debate from a network perspective. *Journal of public administration research and theory*, 13 (2), 193–212.

Volker, L., et al., 2014. The impact of innovative contracting on asset management of public infrastructure networks. *In: Engineering Asset Management*, 665–676. New York City, NY: Springer.

Volker, L., Scharpff, J., and de Weerdt, M. M. (2012). Designing a dynamic network based approach for asset management activities. *In: Proceeding of the 28th Annual ARCOM Conference*, pages 655–664.

Wenzler, I., and Chartier, D., 1999. Why do we bother with games and simulations: An organizational learning perspective. *Simulation & gaming*, 30 (3), 375–384.

Zietlow, G., 2005. Cutting costs and improving quality through performance-based road management and maintenance contracts-the Latin American and OECD experiences. Birmingham: Senior Road Executives Programme, Restructuring Road Management, German Development Cooperation.

Zou, K.H., Tuncali, K., and Silverman, S.G., 2003. Correlation and simple linear regression. *Radiology*, 227 (3), 617–628.