Artificial Intelligence 241 (2016) 1–44

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

# Integrating social power into the decision-making of cognitive agents

Gonçalo Pereira *, Rui Prada, Pedro A. Santos

*INESC-ID and Instituto Superior Técnico and Universidade Técnica de Lisboa, Avenida Professor Cavaco Silva, Taguspark – Edifício IST, 2744-016 Porto Salvo, Portugal*

ARTICLE INFO

*Article history:*
Received 22 May 2015
Received in revised form 20 May 2016
Accepted 10 August 2016
Available online 26 August 2016

*Keywords:*
Social power
Social intelligence
Cognitive architecture
Autonomous agents
Behavior expressiveness

ABSTRACT

Social power is a pervasive feature with acknowledged impact in a multitude of social processes. However, despite its importance, common approaches to social power interactions in multi-agent systems are rather simplistic and lack a full comprehensive view of the processes involved. In this work, we integrated a comprehensive model of social power dynamics into a cognitive agent architecture based on an operationalization of different bases of social power inspired by theoretical background research in social psychology. The model was implemented in an agent framework that was subsequently used to generate the behavior of virtual characters in an interactive virtual environment. We performed a user study to assess users' perceptions of the agents and found evidence supporting both the social power capabilities provided by the model and their value for the creation of believable and interesting scenarios. We expect that these advances and the collected evidence can be used to support the development of agent systems with an enriched capacity for social agent simulation.



## 1. Introduction

Research on intelligent autonomous agents has long focused on developing mechanisms to improve how agents sense, keep a record of and interact with their environment [15,27,70]. As part of this progress, in recent years, there has been increasing interest in social concepts that may contribute to new advances concerning social intelligence and believability in agents [17,24]. The acknowledgment of the importance of these concepts has impacted research in several domains [53]. For instance, agents with a model of cultural dimensions were used to develop an intercultural training tool [55], and the application of social cognitive processes is a key factor in the development of virtual humans to train social skills, such as negotiation, interviewing and leadership [47].

Social power is one of the most pervasive social concepts in human societies because of its function as a *social heuristic* [45] for decision-making. It combines diverse and complex decision-influencing social factors, such as formal/informal norms, resource/action dependencies and social status [16]. The impact of social power may be observed in a multitude of social processes, such as coordination, delegation, cooperation, hierarchy and alliance formation, resource allocation and negotiation [16,58,62]. Reinforcing this idea, it has been argued that power is a cognitive mediator for behavior that is fundamental for emulating many social phenomena that depend on the human social mind [17]. Given the ample impact

* Corresponding author.
*E-mail addresses:* goncalo.pereira@gaips.inesc-id.pt (G. Pereira), rui.prada@tecnico.ulisboa.pt (R. Prada), pedro.santos@tecnico.ulisboa.pt (P.A. Santos).
*URLs:* http://gaips.inesc-id.pt/gpereira (G. Pereira), http://gaips.inesc-id.pt/rprada (R. Prada), http://gaips.inesc-id.pt/psantos (P.A. Santos).

http://dx.doi.org/10.1016/j.artint.2016.08.003

of power in people's decisions, which may be observed in scenarios ranging from simple social interactions [45] to social dilemma situations [3,78], the ability to understand power-based social dynamics and emulate them in cognitive agents is of fundamental importance for improving such agents' interactive capabilities [14].

Although the role of power in agent behavior has previously been explored (e.g., [2,16,40,52]), significant research gaps remain with regard to a comprehensive approach to social power, despite the promise of its use in both inter-agent [54] and human–agent interactions [64]. First, social power may derive from different bases (e.g., reward, coercion, legitimate, expert, or referent [37]) with particular dynamics, but many of these bases of power are frequently not addressed in agent models. In particular, most agent models explore social power dynamics related to legitimacy by using social norms and may include reward and coercion mechanisms, but expert and referent power are rarely considered. Second, even in cases in which we can relate agent models with some of these bases of social power, they are not modeled together as a comprehensive system. There is a lack of agent models that integrate each base of social power and its corresponding underlying factors into the agents' processes of sensing, reasoning and strategic interaction with the environment. We see this as a shortcoming that limits the range of social situations that can be handled by such agent systems for either social simulation or the creation of interactive virtual environments.

Therefore, our motivation is to address this limitation by developing an agent model that integrates comprehensive social power dynamics into the decision mechanisms of a cognitive agent architecture. To achieve this, we first explore the links between different bases of social power and their underlying factors, according to social science research, and propose a formalization of the different bases and their underlying mechanics. This formalization serves as the foundation for the development and implementation of a model that integrates social power processes into the typical agent's decision process. We argue that by using this model, general social power dynamics are granted to the agents, empowering them with intelligent social power behavior that is independent of the context, thereby facilitating the adaptation of the agents to different situations.

Several areas of application can benefit from agents endowed with social power. One is education, in which agents are often used to assist people in learning new skills by presenting area-specific challenges and adapting to the learner's characteristics [41,69]. An example is a leadership training application in which a person training to become a leader would learn to efficiently exercise his/her social power over his/her employees. Another application area is social support, in which agents have been used to support people in coping with difficult situations [79]. For instance, agent-based systems can be applied to aid individuals who are subject to intense social pressure (peer pressure) due to group pressure toward the adoption of unhealthy/negative group norms, which can lead to problems such as violence, bulimia, alcohol abuse or emotional distress [35]. An agent capable of reasoning about the social-power forces involved in social situations could help to analyze sources of pressure and propose strategies for resisting the negative influence of peer pressure. A third application area is entertainment; many games present rich worlds, including increasingly more convincing societies of agents, with the aim of increasing player engagement and "suspension of disbelief" [6]. The central goal for these agents (both allies [25] and adversaries [80]) is to increase their believability by ensuring that they exhibit socially competent interactions and behaviors. For example, in role-playing games, agents that are socially intelligent with regard to social power could be used to create virtual actors that are capable of being either friendly or hostile, depending on the various social concepts that underlie social power. A final example of a potential area of application is social simulation, in which the creation of multi-agent simulations is frequently used to explore human social processes. Social power simulations can be used to study societal behavior in response to changes in the parameters affecting social power dynamics and changes in the structure of the social environment. Such simulations can, for example, be used in training applications to assist health-care workers in coping with the complex social power dynamics involved in a hospital infection control setting [72] or for studies of leadership in small societies [81].

In the case of the application of socially intelligent agents in interactive systems (e.g., training applications, serious games and games for entertainment), the agents should present human-like qualities to improve the human–agent interaction. Humans are intrinsically social beings, and as such, they continuously attempt to understand the meaning of the actions of others. In fact, they attempt to understand actions performed by representations of people, as well (e.g., virtual agents [27]). Consequently, visual and/or intellectual believability are key to the effectiveness of an interactive experience [4]. These traits help users to imagine the experience as real without presenting conflicts that might hinder the "suspension of disbelief" and place the plausibility of the experience at risk. Because social power plays a key role in the cognitive processes that mediate behavior, it is also a key factor for ensuring the believability and effectiveness of an interactive agent system. Hence, to explore the applicability and potential impact in this area of the model proposed herein, we implemented the model in the virtual characters of a game scenario and performed a user study to assess users' perceptions regarding the social power awareness of agents. As a result, we obtained evidence supporting the impact and value of the social power capabilities introduced by the model.

This document is structured as follows. In the section titled Related Work, we review previous relevant contributions regarding social power for agents. Then, we present the main theoretical background research applied in our proposed social power model, followed by the developed model itself and the agent architecture used to integrate it into typical agent processes. Afterward, we describe the implementation of the model and architecture in an agent framework and present the Social Theatre environment that was used to perform the user study to assess the impact of the model. Finally, we draw several conclusions.

## 2. Related work

Before proceeding to the discussion of research addressing the issue of modeling social power, it is important to situate the issue of social power interaction in the context of multi-agent systems. A fundamental work that helps us to do so is that of Castelfranchi on the concept of social action [15]. At its core lies the claim that sociality emerges from agents' goal-oriented minds and their mutual interference in a shared world. The coexistence of multiple agents means that the actions of one affect the goals and plans of the others. To cope with this situation, an agent has the option of either adapting its own goals/plans or changing the goals/plans of the interfering agent. Note that this interference can be either positive, facilitating one's goals, or negative, hindering those same goals. The concept of social action is thus distinguished from the broader concept of action (interaction between an agent and the environment) by the definition that "a social action is an action that deals with another entity as an agent, i.e., as an active, autonomous, goal-oriented entity" [15].

This stance has a strong impact on the conceptualization of diverse aspects of sociality, such as assistance, cooperation and competition. A key element of social action is delegation, which can be summarized as one agent's reliance on another to perform a given action. The precursor to an act of delegation is an agent's recognition of another's action that it likes, values or somehow depends on in order to achieve its goals. However, because the delegated agent is an autonomous and goal-oriented entity, the key problem becomes one of how to induce the delegated agent to perform or adopt the goal of performing the delegated action. It is at this stage of the social action that social power and influence are essential. According to Castelfranchi [15], the power of the delegating agent to influence the delegated agent arises based on the social power of the former over the latter. In effect, this capability of influence is how the delegating agent is able to modify the delegated agent's beliefs and, consequently, its goals and intentions that might lead to the desired behavior. However, this key capability is not possible without the social power to support it, and thus, the modeling of social power is crucial. As acknowledged by Castelfranchi [15], social power that supports influence can originate from various basis, such as authority, sympathy or the exchange of rewards.

### 2.1. Modeling social power

The subject of social power, namely, the representation of power and the formalization of its associated dynamics, has previously been researched from several perspectives. The fundamental differences among the different research approaches lie in the underlying definition of power, i.e., the main focus of the approach being described.

One approach to power is based on autonomy and is rooted in an agent's capability of pursuing its goals without the intervention of other agents. Hexmoor [40] conceptualizes power as a function of personal weights and liberties over preferences. Liberties express freedom/inhibition forces with regard to an agent's decision and can be either internal (e.g., values and emotions) or external (e.g., limitations of the physical context). The weights affect the liberties according to an agent's characteristics of individual rationality (corresponding to a preference for ensuring individual welfare) or social rationality (a preference for social welfare).

Much research is deeply rooted in the seminal work on Dependency Theory by Sichman and colleagues [74,75], which supports a taxonomy of dependence and the fundamental concepts for agent reasoning in groups through the formation of dependence networks. For Carabelea et al. [12], power is grounded on an agent's individual powers, and social power is formalized as originating from the dependence relations of one individual regarding the individual powers of another. These individual powers are generically defined as a combination of the capability and the entitlement (i.e., the right) to do something. Based on this definition, social power is identified in situations in which one agent depends on an individual power of another and both agents are aware of this dependence relation. Castelfranchi et al. [18] differentiate two types of dependence. First, non-social dependence characterizes a dependence relation targeting a resource (e.g., money). Second, social dependence characterizes a dependence relation targeting an action/ability of an agent. However, non-social dependence may give rise to social dependence. If an agent X controls a resource that is needed by another agent Y, then social dependence arises. Agent X controls the resource, and therefore, agent Y depends on agent X's control of that resource. These relations of dependence are the origin of influence interactions between agents, because one agent may promise/threaten another based on such a dependence situation. This notion of social power was also used by Sichman to study coalition formation based on social dependence [73]. Following similar underlying concepts, Castelfranchi [16] developed a concept of social power arising from individual powers that include "internal powers" (e.g., skills) and "external powers" (e.g., resources). Based on this definition, the author explored how social power emerges from the inter-agent dependence of individual powers. The model was further expanded to include other bases of power that emerge from agents' comparisons of individual powers, from the provision of assistance or the exchange of individual powers between agents, and from the combination of individual powers to achieve combined powers. In [7] Brainov and Sandholm relate dependence to utility-based Decision Theory to enable reasoning with regard to power in multi-agent plans. Power is conceptualized based on the possible costs and damages to the agents involved in a given dependence situation that are expected as a result of that situation. These values are defined as the differences in utility between an intended joint plan and a possible alternative plan offered by one of the agents. In this work, power is defined as the capacity to harm another agent, i.e., to execute a plan that can lower the utility of another.

Another approach, adopted by López [52], addresses power from a normative perspective by emphasizing the part played by norms and roles in a group's power structure and individual powers. In this work, López introduced a taxonomy of

powers based on the definition of a normative context, namely, the restrictions or benefits to which an agent in a given group is subject or entitled. These powers are then operationalized in several group processes, such as agent membership in the established normative society, individual norm adoption and inter-agent goal delegation. Also following a normative approach to power, Jones and Sergot [43] presented a logical language for describing the institutionalized powers that arise in norm-governed institutions. At the core of this power conceptualization lies the concept of normative states of affairs that are realized based on the rights and obligations of the entities of a given institution or society. A fundamental issue discussed is the distinction between "permission to do" and "empowerment to do", or institutionalized power. Such a difference might arise in a situation in which an individual has the power (within the institution) to perform an action, but the actual execution of that action might not be permitted (under certain constraints). In the developed framework, power is transferable, by some means of social influence, only when normatively established by the inheritance of institutional power. Influence is defined as the control of the behavior of an institutionally empowered agent. The authors provide several examples of the expressive range of the introduced logical framework in the contexts of ecclesiastical institutions, faculty institutions and typical boss/employee relationships. The depicted situations encompass several social processes, such as delegation and authorization.

Adopting a cultural perspective, Mascarenhas et al. [56] developed the Social Importance Dynamics (SID) Model. It was developed to create agents that exhibit culturally adaptive behavior that is coherent and representative of a (specifically parameterized) underlying culture. The model is based on Kemper's Status-Power Theory [46] because of its focus on daily cultural rituals. More specifically, the status dimension is focused on and modeled as social importance. By modeling this relational characteristic in an agent, the model represents the extent to which that agent is willing to act in the interest of another social entity above its own motivations. The model was implemented in an existing agent architecture [23] following the BDI paradigm, and its impact on the perception, deliberation and planning processes was investigated. Furthermore, to test the expressiveness and practical application of the model, it was used to create synthetic cultures for an intercultural training virtual environment.

A different approach to power relies on power indices, which are frequently used to measure the power of players in voting games. The measurement of power is performed by calculating the importance of a voter in the game. For instance, Bachrach et al. [2] conceptualizes power indices that explore power as a way of measuring the impact exerted by an agent on the outcomes of simple coalition games (i.e., games in which the outcomes can be classified only as successes or failures). The presented power indices measure the criticality of an agent in different possible coalitions.

Hayes et al. [39] assume power to derive from agent interaction through processes similar to those of economic exchanges [63]. Social power is regarded as an exchangeable natural resource in the agents' environment, where it is replenished or decays automatically. For one agent to influence another, it must transfer a given amount of power to the other. In such a situation, the target agent can resist the influencer by transferring a greater amount of power back. In this context, individuals strive to maximize their rewards while lowering their costs for influence.

Although not focused on power, reputation research is related to social power because reputation is a representation of how an individual is regarded in a society. One example of the quantification of reputation by Sabater and Sierra [71] relies on equations used to calculate individual reputations based on a database of individual impressions. The authors also explored a "Social Reputation" metric, which calculates a group-derived type of reputation combined with a type of reputation propagated from other members of the group. The transmission of reputation is therefore a very important mechanism to consider because it may affect the changes in power derived from reputation. This work was subsequently used to investigate social structures by incorporating the reputation dimension in a typical social network analysis.

Two other works that are not directly related to social power nevertheless incorporate certain aspects of the power of interactive virtual agents in the context of influence. From a social power perspective, influence is the exercise of power. Marsella et al. [54] developed the PsychSim agent-based system to simulate social interactions. Agents can affect each other's beliefs and subsequent actions by means of direct interaction (direct messages) or by performing actions that are observed by other agents. The effectiveness of an influence interaction is determined with regard to five underlying factors extracted from a survey of the social psychology literature. These factors (some directly related to power, e.g., likability) affecting the process of interpersonal influence interactions capture the beliefs, perspectives and history of interactions of all participants. Prada et al. [64] developed the Synthetic Group Dynamics (SGD) model with the intent of improving believability in groups of synthetic characters designed to interact with humans. The strength of an individual's proposed group interaction depends on its position in the group, which, in turn, depends on its social relations with others (i.e., social attraction and social influence). The dynamics of these social relations can be related to different bases of power and are operationalized in group interactions between the virtual agents and the player.

### *2.2. Discussion*

To facilitate analysis of the surveyed work, we introduce a set of features identified as important for modeling social-power-intelligent agents. When describing the different works above, we focused on the highlights of each, but these features will help us to perform a more complete and in-depth analysis. The features considered focus on two fundamental areas of interest, namely, the representation of social power in multi-agent systems and its operationalization, which are crucial for the modeling of social power in agents. A comparison of the surveyed works with regard to these features is presented in Table 1.

**Table 1**
Comparison of various approaches to social power. A table cell colored in green (or light gray) indicates that the corresponding model satisfies the requirements of the corresponding dimension, and a table cell colored in red (or dark gray) indicates that it does not.

| Authors | Social Power | Bases of Social Power | Effects of Social Power | Social Science Background | Susceptibility & Influence | Social Power Strategies | Quantification |
|---|---|---|---|---|---|---|---|
| Hexmoor [40] | | | | | | | |
| Carabelea et al. [12] | | | | | | | |
| Sichman et al. [74,75] | | | | | | | |
| Castelfranchi et al. [18] | | | | | | | |
| Castelfranchi [16] | | | | | | | |
| Brainov et al. [7] | | | | | | | |
| López [52] | | | | | | | |
| Jones and Sergot [43] | | | | | | | |
| Mascarenhas et al. [56] | | | | | | | |
| Bachrach et al. [2] | | | | | | | |
| Hayes et al. [39] | | | | | | | |
| Sabater et al. [71] | | | | | | | |
| Marsella et al. [54] | | | | | | | |
| Prada et al. [64] | | | | | | | |

The first main set of features represents **Properties of Social Power** and addresses the conceptualization of social power in terms of its origin, its definition and the effects associated with its use. The considered dimensions are as follows:

**Social Power:** a specific representation or a conceptualization of power that accounts for the social forces derived from an agent's relationship with its social environment (beyond individual powers).

**Bases of Social Power:** the different individual resources (e.g., skills) or social resources (e.g., social norms) from which social power can originate. This dimension analyzes the representation of different bases of social power, with particular focus on the bases of the social power taxonomy introduced by French and Raven [37].

**Effects of Social Power:** the important potential effects of social power on the beliefs and plans of or social relations between individuals in a society. This dimension analyzes the representation of these effects.

**Social Science Background:** whether a model is based, inspired or closely related to any psychological/sociological theoretical background research (behavior models, studies, etc.) that support the way in which social power is conceptualized.

The second main topic is **Operationalization of Social Power** and addresses the modeling of the mechanisms through which social power can be manipulated:

**Susceptibility & Influence:** the existence of a mechanism through which individuals can be affected or affect others at the level of their beliefs, plans or goals through representations of social power.

**Social Power Strategies:** the modeling of strategies that can be used to emphasize an agent's social power (e.g., a social attraction relation or a threat).

**Quantification:** a numerical representation of social power that can be used to operationalize it.

Previous research focusing on social power has presented and explored several possible representations, but these representations also have several shortcomings with regard to the operationalization of social power in intelligent agent systems. One of the identified problems is the typical high level of abstraction of the developed models, which limits the range of social behaviors that can be simulated using those models [2,16,39,40,43,56]. Although several works have addressed bases of social power to some level, in most cases, only one or two of the bases are addressed [7,12,18,43,71,74,75]. Furthermore, some researchers have related their conceptualizations to specific social constructs, but in the end, there is no detailed link between these constructs and their operationalization [12,16,56]. Furthermore, in some studies [12,71], the contexts for which the models were developed are more adequate for a structural analysis at the society level than for an agent behavioral model that operates in a non-deterministic and dynamic environment.

Another problem with current models [16,43,52] is their lack of mechanisms for integrating social powers into the agents' processes. For instance, how do conflicting social powers interfere with each other? Even when certain basic mechanisms are provided, they are typically applied to different social powers in a similar manner, although, in fact, different social powers have different dynamics [39,40,52]. As a result, the behavioral expressiveness potential of existing models either is very limited or has not been explored.

Research related to social power from a social influence perspective is more closely related to agent applications employing social power. However, such studies only superficially address the processes associated with social power in interactions, their different underlying factors and their interdependencies (a taxonomy of social powers) [64]. Because social power is not the focus, these systems specifically lack social power operationalization mechanisms that can emulate different modes of social power utilization and their feedback effects on agents [54].

Additionally, in the surveyed research, there is rarely any reference to the extensive background research on social power that has been performed in the field of social sciences, which could provide a stronger theoretical basis for the developed social power models. The works of Marsella et al. [54] and Prada et al. [64] do acknowledge this theoretical background but are unrelated to social power. Two notable exceptions are the work of Castelfranchi [16], in which his model is compared with several social power theories from the social sciences, and that of Mascarenhas et al. [56], which present the SDI model inspired by the Status-Power Theory [46]. Similarly, two of the least widely addressed topics are the effects of social power and the utilization of social power strategies. Several works address the effects of exerting social power. The cost/damage mechanism presented by Brainov et al. [7] and even the mechanism of social power transfer as presented by Hayes et al. [39] capture some of the consequences of using social power. However, if such an analysis is not supported by a more in-depth evaluation of the origin of the costs, then it provides little insight into the real effects in terms of changes to the underlying factors of social power, such as trust or social attraction. As a result, there has been little focus on understanding how different social power factors impact agents' social reality, which could be used to support simulations with evolving social power dynamics. Social power strategies are, to some extent, addressed in the work of Sichman et al. [74,75], Carabelea et al. [12] and Castelfranchi [16], in which a dependence analysis is used for the selection of partners or for determining group membership. Nonetheless, there is still a gap in current research on this topic because such approaches only indirectly address the wielding of social power as intended by a social power strategy. Regarding the quantification dimension, several works [40,54,56] can be used as an inspiration for quantifying social power.

Furthermore, there has been almost no focus on understanding the link between social power models and possible agent cognitive processes (as identified in theoretical background research) that could be used to operationalize these models in cognitive agent applications. Given these shortcomings, we believe that it is essential to conceptualize social power processes and integrate them with the typical cognitive processes of intelligent agent. This will enable coherent social interactions in which agents can strategically use their social power over other agents and users in dynamic and uncertain environments.

## 3. Theoretical background

We argue that during the process of designing social power mechanisms, it is of fundamental importance to support modeling decisions based on the background provided by social psychology research. There is an extensive body of literature and research on social power that can provide the desired insights. By guiding the conceptualization process with insights from theoretical background research, we leverage the knowledge of experts on social power and, in this way, provide a more solid basis for creating a computational model.

We begin by presenting and discussing various approaches to social power in social psychology and then follow with a discussion exploring how the surveyed research inspired the core of a social power model for agents.

### 3.1. Social power in social psychology

Over the past century, much research has been performed, comprising diverse attempts and perspectives, toward an understanding of social power [34]. As such, social power theories may be categorized from various perspectives [9,34,63]. However, for an attempt to model social power in a cognitive agent architecture, a particularly beneficial categorization is one that differentiates theories according to the stage of the social power interaction process to which they apply. Fiske and Berdhal categorized several theories of social power by placing them in a "causal continuum from source to effect" [34]. In Fig. 1, we present this categorization alongside the key theories representative of each associated category, which we will briefly discuss here. These theories are among the most relevant theories arising from studies of social power in the social sciences.

According to this categorization, social power may be modeled at three distinct levels:

**Outcome Control** Social power is conceptualized as a structural property of social relations.
**Potential for Influence** Social power is conceptualized as the capacity to induce forces that cause influence.
**Influence** Social power is conceptualized based on its effects.

Emerson [28] introduced one of the most influential theories of social power in which social power is modeled from the perspective of **Outcome Control**. According to Emerson, social power is a property of a social relation and cannot exist outside of the conceptualization of a dyadic relationship.[1] Social power is thus defined in terms of dependence: "power

[1] This dyadic relationship need not necessarily occur only between two persons. In fact, it can be any one of the following: a person–person relationship, a group-person relationship or a group–group relationship.

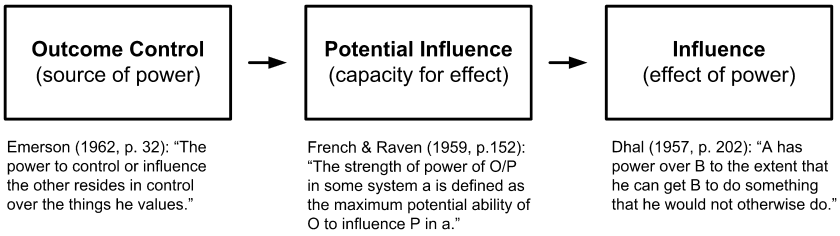


**Fig. 1.** The social power "causal continuum from source to effect" [34] presented by Fiske and Berdhal [34] facilitates an understanding of the different approaches used to study social power. References: Emerson [28]; French & Raven [37]; Dhal [20].

resides implicitly in the other's dependency" [28]. More specifically, the social power of one actor, A, over another, B, is equal to the dependence of B on A. The definition of dependence between two actors is based on two characteristics: "The dependence of actor A upon actor B is (1) directly proportional to A's motivational investment in goals mediated by B, and (2) inversely proportional to the availability of those goals to A outside of the A–B relation" [28].

Another of the most influential [9,28,34,35] and highly regarded bodies of research on social power is the work of French and Raven [37,68], which is representative of the modeling of social power in terms of the **Potential for Influence**. In this approach, social power is defined as a result of diverse psychological forces that have the potential to direct a person's behavior. Lewin and Cartwright were two of the most important proponents of conceptualizing social power as a potential [13,51]. Based on this notion, the specificity of social power is emphasized. An individual may have social power over a certain aspect of another (e.g., how an employee performs his tasks) but none over other aspects (e.g., how friendly an employee is to other people). The main purpose of such a study is to identify the major bases of social power in dyadic relations such that their different effects may be conceptually differentiated:

**Reward** power stems from the ability of one individual to mediate rewards for another. An example of this base of social power is as follows: a factory manager tells a worker that if he increases production, then he will receive a bonus payment.

**Coercive** power stems from the ability of one individual to mediate punishments for another. An example of this base of social power is as follows: a factory manager tells a worker that if he does not increase production, then he will be fired.

**Legitimate** power stems from internalized values that give one individual the authority to influence another. An example of this base of social power is as follows: in a family, a parent instructs a child to be home before midnight.

**Referent** power stems from the identification of one individual with another. An example of this base of social power is as follows: a person asks a friend for help in studying for an upcoming exam.

**Expert** power stems from one individual's perception of another's higher knowledge. An example of this base of social power is as follows: a physician instructs a patient to follow a given medical prescription.

In addition to this work on the bases of social power, Raven also introduced the Power/Interaction Model [66,68], which explores the links between the bases of social power and their operationalization in an influence interaction. According to the authors, there are three aspects that have a strong impact on an influence attempt: the Mode of Influence, the Preparatory Devices and the Motivation for Influence. The Mode of Influence concerns the fact that the effectiveness of an influence attempt appears to be a function of both the bases of power and the means by which that power is exerted. The manner of exertion, can be, for example, loud, forceful, threatening, sarcastic, soft, friendly, or humoring. The Preparatory Devices are the enablers or emphasizers of a given base of social power; for example, "to use coercion, it is sometimes necessary to first make the target realize that the agent has both the means and the will to follow through on the threat" [66]. The Motivation for Influence comprises the various motivational factors that might influence the choice of which bases of social power are to be used in an influence attempt. For example, such factors might include the attainment of extrinsic goals, the desire to adhere to social norms or even a need for status. The Power/Interaction Model also defines the main stages underlying the cognitive process of an influence interaction. Because there are always two participants in an influence interaction, it is presented from both perspectives (those of the influencing and the influenced persons). This model connects the conceptualization of the cognitive process of an influence interaction with the previously introduced bases of social power.

Another work in the "Potential for Influence" category, which has recently been used as the basis for the creation of culturally appropriate agents [56], is the Status-Power Theory presented by Kemper [46]. This theory focuses on the study of the compliance of an individual with the (implicit or explicit) request of another through status or power. Status represents an abstract capacity (with no specific capacity determinants) to gain compliance from another individual through that individual's voluntary action. By contrast, power is conceptualized as the ability to gain compliance through non-voluntary means (e.g., coercion, threats, force), i.e., the capacity to compel non-voluntary compliance regardless of any personal resis-

tance. Kemper's power/status dynamics are related to the generic[2] ecology of the social power structure among individuals and not to the different resources that give rise to that social power. A status claim is made by an individual attempting to influence another; if that fails (status is not conferred), an individual can use his/her power to require the other to confer status to him/her.

In another widely referenced work, Dhal [20] introduced a different perspective on social power, modeling it based on empirical data concerning its effects. This work is representative of the modeling of social power in terms of **Influence**. Social power is seen as a person's capacity to cause another to do something that he/she would not otherwise do. More specifically, an individual's amount of social power over a target individual performing an action, when the first exerts some means of influence, depends on the effect of that means of influence on the target's decision to perform the action. Thus, social power is measured as the difference between the probability of the action being performed when the means of influence is applied and the probability of it being performed when the means of influence is not applied.

Based on this simple definition, the author compared the amounts of social power present in a society of actors. Two important properties discussed were the possibility of ranking the different actors and the transitiveness of the relationships of social power. This conceptualization of social power was also used to rank senators' influence on senate decisions based on previously collected data (from 1946–1954) concerning two types of actions: those "working for" and "working against" a bill on foreign and tax policy.

However, we wish to model social power and its effects based on atomic factors and to simulate social power interactions based on those factors. As such, approaches such as Dhal's, which define social power based on observations of its effects and do not consider the origins or underlying factors of that social power, are not adequate for our modeling purposes. Dhal's approach is especially useful for post-interaction analyses (social power ranking) in which all empirical data on the interactions of interest are available.

### *3.2. From social psychology to agent modeling*

In line with Emerson's conclusion that "power is a property of the social relation" [28], we argue that the cornerstone of and most pervasive factor influencing social power is the existence of some type of valued relationship. If a given relationship is valued by one of its members, then that member depends on the relationship and, consequently, depends on the other member. This valuation confers social power to the other member because the preservation of the relationship ultimately relies as much on the actions of one member as those of the other.[3] Indeed, Emerson's conceptualization of social power in terms of dependence captures the possibility of capitalizing on these valued relationships in a simple framework. Dependence is thus a very useful abstraction that enables us to perform structural analyses of agent societies [5,7,74] and to operationalize various concepts of social power [12,16]. However, the same abstraction that helps us to pinpoint the fundamentals amid a multitude of definitions of social power means that this conceptualization is also too abstract to, by itself, enable the creation of intelligent agents that are capable of strategic social power interactions. The reason for this is that the abstraction eliminates many of the subtleties that differentiate and shape social power in a society (e.g., the leveraging of attraction or normative relations associated with different effects). Emerson's theory is perfectly adequate for the aforementioned purposes, but for an agent society in which different types of social relations have different effects and the social structure itself is continuously changing, a semantically stronger framework of social power is needed.

As Dhal [20] argued, saying that "A has power over B – is not very interesting, informative, or even accurate". It is at this point that more characteristics of social power must be defined to create a semantically stronger framework, and thus, the research of French and Raven [37,66,68] becomes especially relevant. French and Raven present diverse contributions related to differentiating and defining categories of bases of social power, the associated means of influence, their range and certain associated effects. It is also at this stage of the discussion that the fundamental difference between the repeatedly mentioned bases of social power and their actual utilization is exposed. As Dhal accurately put it, "the base of an actor's power consists of all the resources – opportunities, acts, objects, etc. – that he can exploit in order to affect the behavior of another" [20, p. 203]; however, "the base is inert, passive. It must be exploited in some fashion if the behavior of others is to be altered" [20, p. 203].

Even though at first glance, these theories may seem disjunctive and incompatible, we argue that this is not the case and that, in fact, the theories may be combined into a single functional framework. Emerson's [28] dependence may be seen as the atomic element of social power, and French and Raven's bases of social power may be seen as broad, distinct (in source, effect, etc.) categories of prototypical types of dependences in social relationships. A mapping of the different bases of social power to a conceptualization of dependence as social power is easily recognized in the work of Koslowsky and Schwarzwald [48], which establishes an interaction model based on French and Raven's bases of social power [37], and also in Raven's Power/Interaction Model [66]. In that work, the "motivational investment" of Dependency Theory is seen as the "motivation for complying" and the availability of dependence only depends on the number of similar relations in a given society.

Based on this, we may then associate different types of dependence relationships with the specific characteristics of each base of social power as defined by French and Raven [37] and, in this way, combine the fundamental contributions of

[2] All determinants are characterized in the status and power dimensions.

[3] A mutually valued relationship, for instance, a relation of mutual attraction, confers social power to both actors.

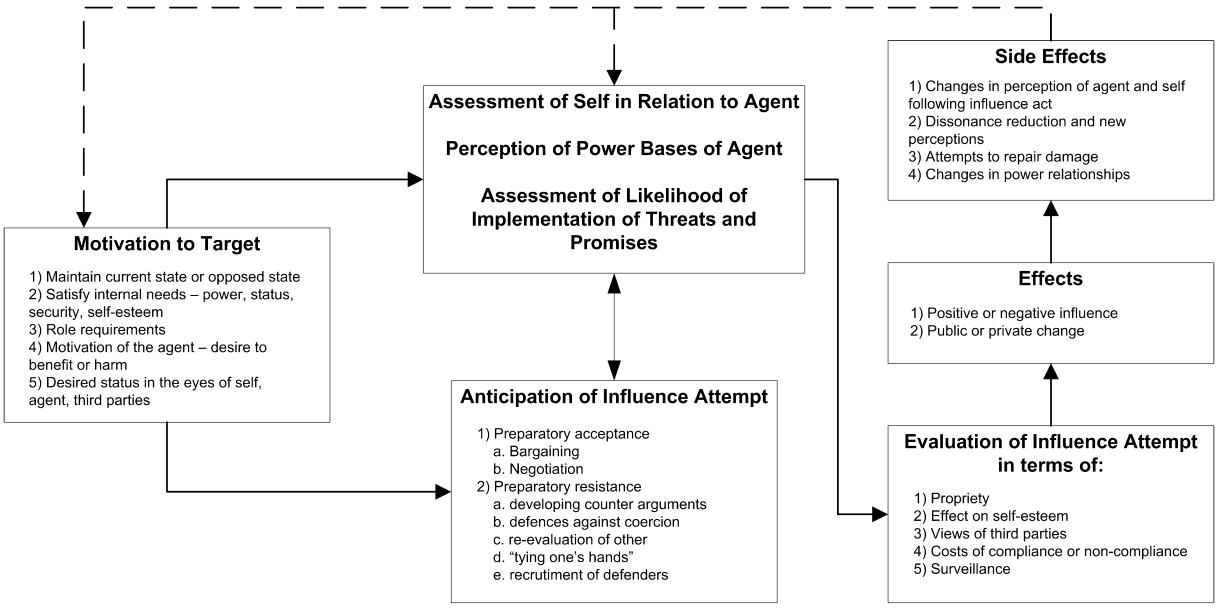


**Fig. 2.** The cognitive process of the influenced agent in a social power interaction in the Power/Interaction Model [66].

both theories for the operationalization of social power interactions. Interestingly, the authors have already provided various aspects of their fundamental interrelatedness. One example of this is the connection established by Emerson [28] between his conceptualization of social power, the concept of group norms as prescriptions of behavior for all group members and the concept of roles as prescriptions of behavior for specific group members assuming a particular role. This link perfectly maps to the base of legitimate social power as described by French and Raven. Another example is French and Raven's [37] discussion of "level observability", which is essentially the dependence of the effectiveness of a social power on the motivational conditions that support it (e.g., a worker in a factory can have different productivities depending on whether the boss is present or absent).

Additionally, Raven [66] has introduced another fundamental contribution to the development of a computational model for social-power-based interactions. Raven introduced the Power/Interaction Model, which provides a conceptual framework exposing and discussing the main social-power-related cognitive processes of both participants in an influence interaction (see Fig. 2). This framework provides invaluable insight for the modeling and operationalization of social power interactions in an agent architecture.

When comparing French and Raven's theory [37] with Kemper's Status-Power Theory [46], we can observe a difference in the levels of analysis of the different types of social power. In the Status-Power Theory, individuals' interactions are explained at the social level and generically include diverse social forces or pressures derived from diverse social constructs that operate on the individual. This conceptualization of power characterizes it as an individual's personal capabilities that he/she can use to gain compliance and possibly status, which will then result in more voluntary compliance in future interactions. From the perspective of French and Raven, the status/conferral dynamics resulting from the use of power consist of different effects derived from social-power-based interactions that affect the social power structure (e.g., increased liking or decreased perceived expertise) and, therefore, the subsequent potential power that individuals possess in the society and can use to exert influence (claim status in subsequent interactions). Further evidence of the link between these theories is the conceptualization of the different forces (personal and social) acting on an individual, which is present in both theories. A relation between the status and power dimensions and the taxonomy of social powers can easily be identified. For instance, status has a clear relation with the legitimate (recognition of norms) and expert (acknowledgment of known expertise credentials) types of social powers, whereas the power dimension can easily be related to the coercive dimension (explicit utilization of coercion).[4]

Although French and Raven present a simple and operational taxonomy of the constituents of social power, their dynamics are not explored in a similarly systematic way, which makes it difficult to understand any basic underlying mechanism of social power dynamics, namely, a type of interaction and the associated effects, in terms of their utilization. By contrast, Kemper's perspective offers an interesting contribution to attempts to model social power dynamics because it provides a generic social power dynamics mechanism that can be applied to the diverse constituents of Raven's conceptualization of social power. Based on Kemper's perspective regarding the claiming/conferral of power/status, the relation between the

[4] Based on personal communication and correspondence with Theodore D. Kemper on the topic of "Status-Power Theory and the French and Raven Bases of Power", from February to May 2014.

utilization of social power (either through a strategy[5] or not) and the consequent effects can also be easily formulated. For example, in an expert power situation, if a given actor uses an expertise strategy ("You should do this because I have more expertise on this topic"), then that actor is explicitly making a claim of higher expert power. If the target accepts this claim, then the target is conferring a higher expert power to the actor and the claim becomes permanent. If the claim is not accepted, then this means that the target confers less power than the actor expected. Similarly, even in the context of a non-strategic interaction (a simple request to perform a given task without explicit support for that request), if the target recognizes the actor to have higher expert power (confers higher expert power to the actor), then the actor's claim is simply met. In the case of a refusal, the target confers a lower expert social power to the actor. Similar reasoning can be applied to various factors that underlie the bases of social power as introduced by French and Raven. Finally, by applying Kemper's dynamics-based reasoning framework to the strategic utilization of French and Raven's bases of social power, we can actually model a wider range of social power strategies than those considered in Kemper's model. This is possible because the claims made by an individual can be focused on strategies other than coercion, such as promises, expertise emphasis, etc.

### *3.3. Supporting individualized agent behavior*

The individual differences among people are factors in determining the diversity of behaviors observed in human societies. To capture this dimension of distinct behavior in humans, in this work, we also strive to create agents that are able to exhibit individualized behaviors. However, to achieve this, we must understand which of the factors that impact individual differences are related to social power dynamics, how they affect these dynamics and how they can be integrated into a conceptualization of social power. From the theoretical background survey, two dimensions stand out as being intimately related to social power dynamics: motivations and personality [66].

Motivation can be described as the driving force(s) behind human behavior that causes people to take and sustain action. Although motivations can vary significantly with individual context, there are certain motivational factors that are present in a wide variety of contexts and have therefore been identified as more pervasive and included in theories of motivation.

Of the many theories of human motivation, one of the most widely used and one that is most strongly related to social power is McClelland's Human Motivation Theory [57]. This theory uses three fundamental needs to explain people's actions in an organizational context:

**Achievement** – the need to set and accomplish goals and to take calculated risks to accomplish those goals.
**Affiliation** – the need to be liked, or the need to belong to the group.
**Power** – the need to control and influence others, or to enjoy status and recognition.

The theory states that all people have these three fundamental motivators, regardless of other individual characteristics, such as gender, culture, and age. By analyzing these dimensions, we can easily relate them to some of the bases of social power (e.g., we can relate affiliation to referent power) and to the motivational cognitive processes in the Power/Interaction Model.

Personality is, by definition, related to the characterization of individual differences by relating them to patterns of behavior. There are many theories that provide various alternative approaches to modeling personality, but one of the most widely acknowledged is Costa and McCrae's five-factor model (or Big 5) [19]. This model characterizes individual differences based on five dimensions of personality traits. A personality trait defines stable personal characteristics that are associated with a specific pattern of behavior. The five factors identified in the model are as follows:

**Openness to experience:** inventive/curious versus consistent/cautious.
**Conscientiousness:** efficient/organized versus easy-going/careless.
**Extraversion:** outgoing/energetic versus solitary/reserved.
**Agreeableness:** friendly/compassionate versus analytical/detached.
**Neuroticism:** sensitive/nervous versus secure/confident.

Even with a clear representation of different patterns of behavior, we still need to understand how these patterns relate to social power. This gap has been addressed in the work of Karkoulian et al. [44], which identifies several correlations between the characteristics of the five-factor model and preference for the utilization of different bases of social power according to French and Raven's taxonomy.

Modeling individual differences between agents based on these theoretical background theories has two additional advantages beyond enabling the creation of individualized agent behavior. First, it helps to reduce the number of free variables in the model that are related exclusively to the model, thereby facilitating the authoring process and the creation of agent simulations. A second advantage is that linking the agent model to these theories enables us to take advantage of existing questionnaires to assess people's ratings according to these models and create simulations using real data.

[5] Power strategies are related to "Preparatory Devices" in the Power/Interaction Model [66].

### 3.4. Fundamental notions and interaction process

Based on the discussion presented in Section 3.1, we conceptualize social power as a potential force [34]. As such, the working definition we use is adapted from Lewin's [51] and Cartwright's [13] definitions of power:

> *"The **Social Power** of A over T regarding a possible behavior of T is the resultant potential force that A can induce on T toward the realization of that behavior."*

Based on Lewin's definition, we characterize power as a potential force that can be induced but might or might not be used. Based on Cartwright's definition, we characterize the specificity of power toward a particular behavior. Finally, we also apply the concept of "resultant" to the potential force characterization to reflect the integrative nature of social power with regard to its different origins from diverse bases. Considering this definition of social power, we define influence as follows:

> *"The **Influence** of A over T regarding a possible behavior of T is the effective utilization of the resultant potential social-power force wielded by A to induce T to perform a specific desired behavior."*

Regardless of the links that exist between different categories of social power (Section 3.2), by using this definition, we situate our approach within the Potential for Influence category (see Fig. 1). This approach enables us to readily account for the dissonances in social power (believed and actual) that agents might encounter in influence interactions. In considering diverse bases of social power, we support the flexibility of the definition to express many different social power situations, thereby avoiding the lack of expressiveness that can result from oversimplification. Furthermore, by endowing social power with specificity toward particular behaviors, we are able to account for the particular dynamics of each base of social power in the social power ecology of a given social environment.

According to Lakoff and Johnson [50], this definition of social power as a potential force can be regarded as a conceptual metaphor. Such a metaphor is very useful because it helps us to define one conceptual domain based on the understanding of another. However, one must be aware that the value of this approach is limited by the possibility of inconsistencies between the two conceptual domains or imperfections in the conceptual mapping. Inconsistencies in the metaphor can cause us to overlook valuable aspects of the concept that we are attempting to model. For instance, in the metaphor "argument is war", a very strong focus might be placed on the "fighting" aspect of an argument, while the potential "cooperative" aspect is completely obfuscated. The imperfections of such a conceptual mapping are related to details of the domain to be defined that simply are not mappable to the metaphorical domain. As such, we must not use these metaphors as literal descriptions of reality but rather as structuring element, while still retaining the complexity and specificity of the concept we are trying to conceptualize.

Before the developed model of social power is introduced, it is important to clearly understand the elements participating in a social power interaction process and to understand the perspectives of these different elements regarding the interaction. The dynamics of the social power interaction process vary depending on how such social power is perceived and leveraged in different situations. To analyze this, we need to consider three core elements that are present in all social power interactions:

**Actor (A):** an individual who exerts social power over the behavior of another.
**Target (T):** the individual whose behaviors are affected by the actor's social power.
**Behavior (C):** the behavior executed by the target in a given interaction. Note that this behavior may be extended from the concept of (physical) action to include other types of changes, such as changes in beliefs, without losing any behavioral expressiveness. For example, in the case of a social power interaction acting on a belief regarding the position X of an object Y, the corresponding action can be to "update the belief that object Y is at position X".

In any social-power-based interaction, these elements are related to the target's reasoning about the actor's social power regarding the execution of a given behavior. However, the dynamics of a social power interaction vary significantly depending on the awareness of the participating individuals, which individual initiates the interaction and the manner in which the power is invoked.

The simplest situation is one in which the actor merely exists in the context of the target and the target must consider the social power relations between the two before performing an action. In this case, the target's personally motivated action is the behavior being evaluated and the social power is invoked only implicitly because there is no direct communication between the participants. Consider a situation in which the actor is a police officer standing on a sidewalk near a book shop. Additionally, it is forbidden to park cars in front of the shop, and the target is a driver intending to park near the book shop despite his awareness that it is forbidden to park there. The simple presence of the actor will probably dissuade the target from parking near the book shop. If he does park there, the police officer will probably punish him with a fine for illegal parking. Here, although no explicit command is issued by the actor, the target is still reasoning about the social powers of other individuals in the area (in this case, the actor/police officer) regarding his own actions without being directly prompted by them (the actor). Interestingly, this reasoning process of the target occurs regardless of the actor's awareness, which depends on the police officer's intentions. In a case in which the actor is aware, the police officer would be there

precisely to dissuade illegal parking. Alternatively, the police officer could also be simply standing on the sidewalk taking a break from work, with no intention of dissuading illegal parking; however, the effect on the target would be the same. Finally, we note that the context in which the police officer is present might be a simple mental context for the target, i.e., the target could simply imagine the possibility of the police officer's presence (for instance, based on previous experiences) and be influenced in a similar manner.

A different situation occurs when the interaction is initiated not by the behavior of the target but by a command from the actor directing the target to perform a given behavior. In this case, both actor and target are always aware of the social-power-based interaction, but the way in which social power is invoked can still be implicit or explicit. As an example, imagine a situation in which an actor asks a target to open a door. If both individuals share the same belief that the actor is entitled to be in that room, then the target will comply without asking for any type of credentials. This is a case in which social power is invoked implicitly. However, if the target does not share this belief, then he/she will probably[6] not comply with the request to open the door. The actor could address this belief mismatch by again requesting the same behavior from the target, accompanied by the presentation of some type of certification to communicate his legitimacy (the basis of his social power) to enter the room. The latter scenario, in which a social power dispute arises from the incompatible beliefs of the participants, represents a case in which social power is explicitly invoked. This assertion is related to the use of social power strategies, which are employed by an individual to emphasize his/her social power bases (in this case, his/her certification) to leverage his/her social power. This and other more complex strategies (e.g., assertion, deception, threat and manipulation [29,30]) are commonly used in interpersonal interactions.

In the presented types of interactions, there is the common element that the target always performs social-power-based reasoning[7] with regard to a behavior before deciding whether to comply with it. Note that this is independent of the individual who initiated the interaction, and although the actor might also perform social-power-based reasoning before commanding the target to perform some behavior, this is not always the case. In the first type of interaction, the actor might not even be aware of the specific social power situation.

Another important aspect of a social-power-based interaction that could extend this interaction process is the presence of multiple influencing agents (multiple actors or supporters for the actor). It has been acknowledged in the literature [45] that the presence of other individuals (beyond the actor and target) can either favor or hinder the influence being exerted, thus effectively impacting the influence process. The impact can originate from different causes. An example of this is related to personal image management [16]. For instance, imagine a situation in which an adolescent smokes when hanging out with his friends as a symbol of his adulthood. He does this to avoid the negative image impact that it could have on his parents, who strongly advocate against smoking, and if, for some reason (e.g., a graduation celebration), his parents and friends were simultaneously present, the adolescent would be strongly influenced not to smoke by the mere presence of his parents. These multiple influence dynamics are an interesting topic of research, but in this work, we focus on the fundamental influence dynamics, i.e., those that occur at a dyadic level. This would appear to limit the types of factors that can be used in modeling social power to those that are strictly dyadic in their conceptualization. However, this is not the case because, for example, group-level concepts such as social groups and social norms can also be considered by individuals when reasoning about dyadic interactions.

## 4. SAPIENT: a model for agent social power

In this section, we present the developed model of social power: SAPIENT – SociAl-Power-Intelligent agENTs. In the basic conceptualization of social power that underlies this model, social power is regarded as a potential force [13,51,37] affecting an agent's decision (see Section 3.4). Additionally, to address the limitations of conceptual metaphors, as discussed by Lakoff and Johnson [50], we must also capture the complexity and specificity of social power. For this purpose, we model a broad range of bases of social power following the work of French and Raven [66,68]. In doing so, we increase the expressiveness of our model by differentiating among different bases of power and their underlying factors, thereby creating distinct social power dynamics. The reason for using this taxonomy of social powers is that it is one of the most influential and highly regarded theories of social power [9,28,34,35], which has been repeatedly scrutinized over the years [42,48,49,60,67,77], and thus provides a strong basis for our work. We begin by introducing basic definitions and mechanisms, followed by the conceptualization of the different bases of social powers and a discussion of their behavioral expressiveness.

### 4.1. Initial considerations

The goal of the social power model that we will present in this section is to enable the representation and manipulation of social power concepts in an agent system independently of the preferred type of agent implementation. For example, an implementation might rely on a utility-based system in which the values of most variables are defined by attempting to

[6] The belief regarding entitlement is relevant only to the base of legitimate social power; other bases could also be at play, and thus, even if the actor does not have permission to be present, the target could still open the door. A base of social power categorizes the type of attributes (social resources) that confer a given social power to an individual [37]. For instance, bases of legitimate social power include both formal and informal norms, contracts and interpersonal commitments [12].

[7] Even if at an unconscious level.

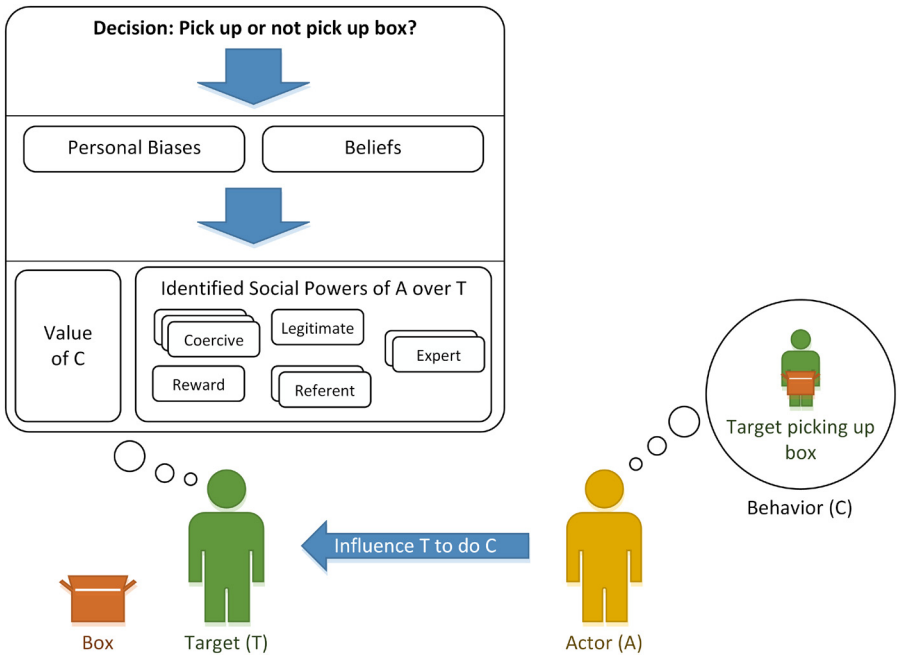


**Fig. 3.** Diagram representing the components of and context for the social power decision mechanism. Note that if the actor intentionally influences the target, it does so with information resulting from analyzing the target's perspective but based on the actor's beliefs regarding the target.

quantify them based on other values or characteristics that contribute to their definition. By contrast, in several rule-based systems, the variables are simply set to represent the desired scenario and then adjusted to correctly match the desired behaviors. As such, throughout the description of the model in this section, we will introduce several variables without specifying their ranges. Because the ranges of values strongly depend on the type of implementation used and the scenario a researcher intends to model, the task of normalizing and adjusting these ranges of values should be carefully performed when modeling a specific scenario. The normalization is especially important in a context in which different variables are combined to model some integrative concept, such as social power. Nonetheless, to facilitate description of the model, we will characterize each variable as one of the following types:

**Positive/Negative** – can assume a positive, negative or zero value.
**Positive** – can assume only values greater than or equal to zero.
**Probability** – assumes a value in a probability range (e.g., typically represented by $[0, 1] \in \mathbb{R}$).

### *4.2. Social power decision mechanism*

The function of social power as a *social heuristic* is relevant in a context in which an agent must decide among multiple possible actions that it can perform. Typical approaches to such decisions include performing some type of assessment of the value to the agent offered by each of the actions and, for example, choosing the action with the highest value as the one to be performed. Such an assessment can be conducted in different ways, from a directly specified table of values to an algorithm that performs a weighted analysis of several factors that are relevant for the current context. In our approach, such a value assessment is retained, as it plays an important part in the agent's decision, but we do not focus on how this assessment is performed. Instead, we add social power as an essential force that influences the agent's decision such that is coupled to the value assessment but independent of the way in which this assessment is performed. We are therefore interested in how this additional force driving the agent's decision is integrated and conceptualized.

Given the working definition of social power as a potential force, and before clearly defining the different bases of social power used in the model, we must first introduce the basic mechanism governing the target's decision process (see Fig. 3). In an influence situation in which an actor ($A$) exerts an influence over a target ($T$) regarding a behavior $C$ that $A$ wishes $T$ to perform, there are only two possible outcomes: $T$ decides to perform $C$ or $T$ decides not to perform $C$. In the first case, the influence interaction is successful; otherwise, it fails.

In a given social power interaction, the agent may identify several social powers to be simultaneously active. Each identified social power ($p$) represents a single social power originating from a specific base of social power. The set as a whole may contain social powers originating from only one base (e.g., may include several coercions) or may comprise social powers arising from several distinct bases [37]. We denote the set of identified social powers by *IdentifiedSocialPowers* (*ISP*). Each of these powers creates a potential force ($PotentialForce_p$) exerted by the actor on the target, influencing the target's

decision. To model the potential force represented by each different base of social power, two types of factors must be considered (see Formula (1)). First are the personal biases ($PB_{T,p}$) related to the preferences or inherited traits of the agent that affect the associated social power (e.g., a need for affiliation [57]); these should be defined for the agent at authoring time. The factors of the second type are the beliefs ($Be_{T,p}$) that evolve from the agent's interactions in the environment that also affect the different social powers (e.g., a belief regarding another agent's tendency to provide a certain reward). There exist a set of personal biases and another of beliefs that are specific to each base of social power. The function $f$ combines the sets of personal biases and beliefs in different ways depending on the base of social power that is associated with each particular social power that has been identified. As such, $f$ is, in fact, a linear combination of functions that consists of a sub-function for each base of social power considered in the model (e.g., $f_{Coercive}$ and $f_{Expert}$).

$$\begin{aligned} &p \in IdentifiedSocialPowers \\ &PotentialForce_p(T, A, C) = f_p(PB_{T,p}, Be_{T,p}) \end{aligned} \tag{1}$$

A social power potential force ($PotentialForce_p$) assumes a value in a positive/negative range, in which positive values represent a positive influence toward performing $C$ and negative ones represent an influence toward not performing $C$ (as a realization of the concept of negative social power [66]). A value of 0 means that, although identified as a potential influence, the force neither favors nor discourages the performance of $C$. To determine the actor's overall influence on the target's decision resulting from all potential social-power forces, we assume a simple resultant force approach in accordance with the working definitions of social power and influence (see Section 3.4). As represented in Formula (2), this overall influence force is the sum of all of the individual potential forces in $ISP$.

$$InfluenceForce(T, A, C) = \sum_{p \in ISP} PotentialForce_p(T, A, C) \tag{2}$$

In addition to the influence resulting from social power, another force also indisputably acts on agent decisions, namely, the value of the action to be performed. We can better elucidate this additional force by comparing two situations. First, let us consider a situation in which a person asks a friend to temporarily lend him a cellphone. It is reasonable for the friend to do so given the friendship relation and the low loss of value he experiences by lending the cellphone temporarily. In the second situation, the person now asks his friend to buy him a cellphone. In this case, it is similarly reasonable for the friend to refuse given the high loss of value he would experience by buying the cellphone. Generally, value measures the usefulness of an agent performing an action as a function of the agent's motivations and goals.

In this mechanism, we model the value force as *ValueForce* to represent the agent's natural disposition to perform action $C$. This propensity is unrelated to any social power bases. In fact, in the absence of the coupled influence of social power, the agent's decision would be based solely on the propensity represented by *ValueForce*. *ValueForce* assumes a value in a positive/negative range, in which positive values represent a situation favoring the decision to perform $C$, negative values represent a resistance situation favoring the decision not to perform $C$, and 0 indicates indifference. Note that the way in which this value is determined is not our focus because it can be defined in various different ways that are unrelated to the key issue of social power.

Finally, we model the agent's decision to either perform $C$ or not $\neg C$ as a combination of the two major forces identified: the influence of social power and the value of the action. To model the decision, we assume a simple resultant force approach, as represented in Formula (3). The $\alpha$ parameter is used to calibrate the balance between the value force and the influence force, thereby facilitating the adjustment of the weight of each component in each decision. For instance, a value of $\alpha = 1$ would mean that the agent would make its decisions based only on the value force, disregarding any social-power-related biases. This balance value can be determined experimentally or can be defined by means of other complementary models or frameworks.

$$\begin{aligned} &ResultantForce(T, A, C) = \alpha * ValueForce(T, C) \ + (1 - \alpha) * InfluenceForce(T, A, C) \\ &Decision = \begin{cases} Do(T, C), & \text{if } ResultantForce(T, A, C) > 0 \\ \neg Do(T, C), & \text{if } ResultantForce(T, A, C) \leq 0 \end{cases} \end{aligned} \tag{3}$$

If the value is positive, then the target decides to perform $C$ ($Do(T, C)$); if not, then the target decides not to perform $C$ ($\neg Do(T, C)$). Note that this decision formalization accounts for the resistance that an agent can offer to a given prescription of behavior from another agent [37]. This is present at two distinct levels. First, if the value is negative, it represents an opposing force to the influence being exerted. Second, each $PotentialForce_p$ (originating from any base of social power) may be a negative potential force, which is another form of opposition to a decision to perform $C$. For instance, when a person dislikes another, this will be represented by a negative force and influence arising from the referent social power base.

### 4.3. Social power dynamics

In the definition of the agent decision mechanism, we introduced the influence force derived from social power as a combination of all of the diverse social powers that are identified as active in a given situation, each arising from a specific

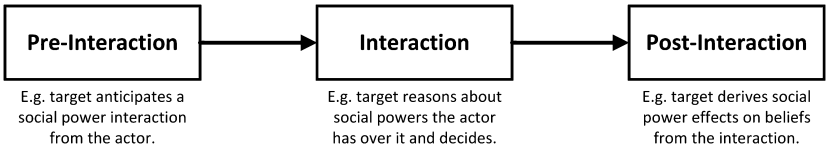


**Fig. 4.** Diagram depicting the main phases of a social power interaction based on the cognitive processes of social power [66].

social power base. In this section, we describe in detail each base of social power according to French and Raven's taxonomy of social powers [37]. The modeled bases of social power are as follows: reward, coercive, legitimate, referent and expert. Note that the descriptions are provided from the perspective of the target of the corresponding influence interaction, i.e., based on the target's characteristics and perceptions. For each base of social power considered in the model, we introduce a conceptualization that enables us to integrate it into the decision mechanism as a function of an agent's personal characteristics and beliefs, as generically introduced in Formula (1). Although all bases of social power are clearly defined, the factors that should be used to conceptualize them are not always so clearly identified by French and Raven [37]. To be able to create a formalism that can subsequently be used in agent simulations, we therefore invoke concepts that were used in the descriptions of the bases of social power and relate them to relevant theoretical background insights into their associated relations.

Note that when describing the bases of social power, it is important to consider how their underlying factors vary to enable the application of our conceptualizations in continuously evolving dynamic environments. To this end, and despite the fact that the dynamics of certain factors (e.g., changes in social group membership) are beyond the scope of this work, we also detail several aspects that are directly related to each base of social power (e.g., a belief regarding an agent's tendency to coerce). These descriptions are presented based on a common temporal evolution of a social power interaction, as illustrated in Fig. 4, and with a focus on the interaction and post-interaction phases because of their contributions to the ecology of social power. A common factor that is used to distinguish the dynamics of each social power is the outcome of the interaction phase. There are only two possible outcomes: the success or failure of the influence interaction.

Additionally, to demonstrate the expressiveness of the presented model, we provide examples of critical incidents involving each base of social power. Once all bases of social power have been introduced, we also present a situation in which most types of social powers are present. Although we considered constructing this example as we described the different bases of social power, in the end, we refrained from doing so. The reason for this is that, to our knowledge, it is not common for almost all types of social powers to be present at the same time, at least, not in a manner to which the reader could easily relate them.

#### 4.3.1. Reward

In an influence situation, a reward social power exists when the target ($T$) recognizes that the actor ($A$) can execute an action ($ra$) that the target positively values in exchange for the execution of the action that the actor intends the target to perform ($C$). Rewards are based on resources such as money or social praise. Note that the set of all possible rewarding actions that the actor can perform for the target are represented by the set $RewardingActions_A$.

A reward social power that an actor exerts over a target that influences the latter to perform an action is a function of the value that the target assigns to the rewarding action that the actor can perform ($Value_{ra}$), the target's achievement motivation ($AchievementM_T$) and the target's perception of the tendency of the actor to perform the rewarding action ($Rewarding_{T,A,ra}$). This description is succinctly modeled by Formula (4).

$$\begin{aligned} \exists_{ra} \in RewardingActions_A \wedge ra \neq C \\ PB_{T,Reward} &= \{AchievementM_T\} \\ Be_{T,Reward} &= \{Rewarding_{T,A,ra}, Value_{ra}\} \\ f_{Reward}(PB_{T,Reward}, Be_{T,Reward}) &= Value_{ra} * AchievementM_T * Rewarding_{T,A,ra} \end{aligned} \tag{4}$$

A reward social power is therefore directly proportional to the value of the rewarding action, the achievement motivation of the target and the target's perception of the rewarding tendency of the actor for that rewarding action. The value of an action is a positive/negative variable, where negative values represent negatively valued actions, positive values represent positively valued actions, and a value of 0 represents an irrelevant action. The achievement motivation is a probability variable, where a value of 0 indicates that the target is not at all achievement motivated and a value of 1 indicates that the target is extremely achievement motivated. The rewarding tendency is also a probability variable, where a value of 0 indicates that such a tendency has never been observed and a value of 1 indicates that it has always been observed.

*Dynamics.* If a reward-based influence interaction is successful, then a commitment is established between the agents. After the interaction phase, the target performs the action that the actor intended the target to perform and then expects the actor to execute the rewarding action. Subsequently, the actor performs the rewarding action for the target.

The personal bias related to achievement and the value of the rewarding action vary only with the internal motivations and goals of the target. By contrast, the rewarding tendency factor evolves based on the target's experience with the actor

agent in accordance with the social power dynamics. If the influence interaction is unsuccessful, an immediate effect is that the actor's perception of the target's susceptibility to reward[8] decreases. If it is successful, then the commitment[9] established between the actor and the target commits the latter to execute the action toward which it was influenced and the actor to execute the expected/promised reward after the target fulfills his/her obligation.

Finally, when the actor provides the expected/promised reward, the target's perception of the rewarding tendency of the actor increases, and the target's liking of the actor also increases.[10] Otherwise, if the reward is not given, the target's perception of the rewarding tendency of the actor decreases.

*Examples of critical incidents.* Imagine a situation in which a factory manager tells a worker that if he increases production, then he will be granted a bonus to his salary. A promise of a higher bonus is expected to exert a stronger reward social power. However, if the worker does not believe that the manager rewards employees with bonuses for increases in their productivity, then there is no reward social power. Similarly, the strength of the social power varies depending on the worker's achievement motivation. If the worker is very weakly achievement motivated, then the promise of a bonus will not have a very strong social power impact, whereas if he is highly achievement motivated, then the social power exerted by the promise of a bonus is much stronger.

A less common situation that can also be modeled using this conceptualization is that of a negative reward social power. Now imagine that instead of a salary bonus, the manager offers to reward the productivity increase with longer breaks during the workday. The intention may be a positive reward, but if the worker has a family and wishes to decrease the total amount of time spent at the factory to increase his family time, then this offer becomes a negative reward because it increases the total amount of time spent at the factory (work time plus break time). Such situations arise when there is a mismatch between the motivations that the actor agent believes the target to have and his/her real motivations.

#### 4.3.2. Coercive

In an influence situation, a coercive social power exists when a target ($T$) recognizes that an actor ($A$) can execute a harmful action ($ca$) if the action underlying the influence interaction ($C$) is not performed by the target. Coercions are based on resources such as money or gossip. Note that the set of all possible harmful actions that the actor can perform against the target are represented by the set $CoerciveActions_A$.

A coercive social power that an actor exerts over a target that influences the latter to perform an action is a function of the value that the target assigns to the coercive action that the actor can perform ($Value_{ca}$), the target's achievement motivation ($AchievementM_T$) and the target's perception of the tendency of the actor to perform the coercive action ($Coercing_{T,A,ca}$). This description is succinctly modeled by Formula (5).

$$\begin{aligned} \exists_{ca} \in CoerciveActions_A \wedge ca \neq C \\ PB_{T,Coercive} &= \{AchievementM_T\} \\ Be_{T,Coercive} &= \{Coercing_{T,A,ca}, Value_{ca}\} \\ f_{Coercive}(PB_{T,Coercive}, Be_{T,Coercive}) &= -Value_{ca} * AchievementM_T * Coercing_{T,A,ca} \end{aligned} \tag{5}$$

A coercive social power is therefore inversely proportional to the value of the coercive action and directly proportional to the target's perception of the coercive tendency of the actor for that coercive action. The value of an action is a positive/negative variable, where negative values represent negatively valued actions, positive values represent positively valued actions, and a value of 0 represents an irrelevant action. The achievement motivation is a probability variable, where a value of 0 indicates that the target is not at all achievement motivated and a value of 1 indicates that the target is extremely achievement motivated. The coercive tendency is also a probability variable, where a value of 0 indicates that such a tendency has never been observed and a value of 1 indicates that it has always been observed.

*Dynamics.* If a coercion-based influence interaction is successful, then a commitment is established between the agents. After the interaction phase, the target performs the action required by the actor and expects the actor's non-execution of the coercive action. If the interaction is unsuccessful, the target expects the actor to execute the coercive action.

The personal bias regarding achievement and the value of the coercive action vary only with the internal preferences, motivations and goals of the target. By contrast, the coercive tendency factor evolves based on the target's experience with the actor agent in accordance with the social power dynamics. Regardless of the success or failure of the interaction, an immediate effect of the use of a coercive social power is that it decreases the target's liking of the actor.[11] However, if the

[8] This susceptibility, although not part of the quantification of the social powers, is important for planning the utilization of social power. During the planning of a given social-power-based interaction, this belief regarding susceptibility weights the expected effectiveness of a rewarding strategy for the planned future interaction.

[9] This agent-specific commitment to action is an underlying factor of a legitimate social power (see Section 4.3.3).

[10] A side effect of using a reward social power [37]. This social relation based on the perception of liking is an underlying factor of a referent social power (see Section 4.3.4).

[11] A side effect of using a coercive social power [37]. This perception of liking social relation is an underlying factor of a referent social power (see Section 4.3.4).

coercion is based on a previously established commitment, then the decrease-of-liking effect is nullified. If the influence interaction is unsuccessful, another immediate effect is that the actor's perception of the target's susceptibility to coercion decreases.[12] If it is successful, the commitment[13] established between the actor and the target commits the latter to execute the action toward which he/she was influenced and the actor to not execute the expected/threatened coercive action after the target fulfills his/her obligation. This commitment is also key to differentiating coercive and reward social powers.

There are three major differences between coercive and reward social powers. First, a coercive social power exists on the basis of the target's perception that the actor performs actions that individuals value negatively instead of actions that individuals value positively. Second, and based on these different perceptions, a coercive social power establishes a commitment for the actor's non-execution of an action, whereas a reward social power establishes a commitment for the actor's execution of an action. Third, the effects on the liking social relation occur in different phases[14] of the social power interaction and are affected by different conditions (e.g., the negative effects of coercion can be nullified). Consequently, although the two types of social power have very similar mechanisms that are based on resources of the same type (but oppositely valued), the distinct dynamics associated with these types of power lead us to model them individually. Additionally, this decision also facilitates the utilization of the model in the context of other research using French and Raven's [37] framework of social power bases. If we were to generalize them under a more abstract concept, we would always be obliged, at some point, to make the distinction between these two types of social powers.

Finally, if the actor performs the expected/threatened coercive action, the target's perception of the coercive tendency of the actor increases. However, if the coercive action is avoided, then the target's perception of the coercive tendency of the actor decreases.

*Examples of critical incidents.* Imagine a situation similar to that used to illustrate reward social powers, but in which the factory manager tells the worker that if he does not increase production, then he will be fired. The strength of the social power varies with the magnitude of the coercion. However, if the worker does not believe that the manager fires employees for failing to increase their productivity, then there is no coercive social power. Similarly, the strength of the social power varies depending on the worker's achievement motivation. If the worker is very weakly achievement motivated (under the assumption that he has a motivation to keep his job), then the threat of being fired might not have a high impact, but if he is highly achievement motivated, then the social power exerted by the threat of being fired is much stronger. As in the case of a reward social power, a negative coercive social power is also possible if there is a mismatch between the motivations that the actor agent believes the target to have and his real motivations.

#### 4.3.3. Legitimate

In an influence situation, a legitimate social power exists when a target ($T$) recognizes his/her duty toward an actor ($A$) to execute the action underlying the influence interaction ($C$) based on a social norm ($n$) established for a group ($g$) to which both agents belong. Note that *SocialGroups* represents the set of all existing social groups to which the target belongs, $Members_g$ represents the set of all members of social group $g$, and $Norms_g$ represents the set of norms of group $g$. Social groups can be created or emerge in societies for several reasons (e.g., common goals) in both formal (e.g., universities) and informal (e.g., groups of friends) contexts [36]. Additionally, the processes that govern the group dynamics of membership, the definitions of roles and the creation/adoption of group norms also follow specific and complex mechanisms. The focus is on modeling the social powers that emerge from such social constructs and not on modeling changes to the group, which will, of course, affect the social powers acting between group members. For instance, if a member of a group abandons his/her membership in that group, then he/she ceases to be under the influence of the social norms adopted by that group. Given the broad scope of possible social norms, our conceptualization is inspired by a basic categorization of norms introduced by Raven [68], which builds upon French and Raven's seminal theory of the bases of social power used to support this model [37]:

**Position:** members of the group are required to obey other members who are in a superior position.
**Reciprocity:** when a member of the group is benefited, that member should feel obligated to reciprocate.
**Equity:** members of the group who have been harmed have a right to compensation.
**Responsibility:** members of the group have some obligation to help other members who are dependent upon them.

Because all of these norms are whole-group norms, in addition to Raven's categories, we also introduce the action commitment social norm to express specific interpersonal commitments between agents within a given social group.

A legitimate norm social power ($n$) is a function of the importance of the relevant social group to the target ($Importance_g$), the intrinsic dutifulness ($DutifulnessM_T$) felt by the target toward the social group, the group norm conformity recognized by the target ($Conf_{T,g}$) and a norm-related bias ($Bias(n)$) derived from the agent's association and experience with the social group. This description is modeled by Formula (6). Note that although some variables from the belief set are not directly used in Formula (6), they are used to formulate the norm bias, which will be further detailed below. Additionally, note that

[12] Similarly to the reward susceptibility, this susceptibility is important for planning the utilization of coercive social power strategies in future interactions.
[13] This agent-specific commitment to action is an underlying factor of a legitimate social power (see Section 4.3.3).
[14] For subsequent interactions in the case of reward and immediately in the case of coercion.

in the context of a given social group, several legitimate social powers can be active at the same time because of different social norms.

$$\begin{aligned} &\exists_g \in SocialGroups \wedge \{A, T\} \subseteq Members_g \wedge n \in Norms_g \wedge NormTypeConditions(n) \\ &PB_{T,Legitimate} = \{DutifulnessM_T\} \\ &Be_{T,Legitimate} = \{Importance_g, Salience_n, Conf_{T,g}, \\ &\qquad HelpfulActions_{g,T}, HarmfulActions_{g,A}, AllActions_g\} \\ &f_{Legitimate}(PB_{T,Legitimate}, Be_{T,Legitimate}) = Importance_g * (DutifulnessM_T + Conf_{T,g}) * Bias(n) \end{aligned} \tag{6}$$

The strength of a legitimate norm social power is therefore directly proportional to all of its components: the importance of the group, the dutifulness felt by the target, the recognized group norm conformity and the norm bias. The group's importance is a positive variable, where values close to 0 represent a weak group association and higher values represent stronger associations with the group. The dutifulness is a probability variable, where a value of 0 indicates that the target has no sense of obligation and a value of 1 indicates that the target has a very high sense of obligation. The group norm conformity component derives from observations of conformity to group norms by other elements of the group and represents a dutifulness bias. For instance, in a setting in which members of the group frequently fail to comply with group norms, the target will feel less duty-bound to comply with a norm that exerts a legitimate social power. The group norm conformity is a probability variable, where a value of 0 indicates that norm conformity has never been observed and a value of 1 indicates that it has always been observed.

The most significant differences between different legitimate social powers result from the type of norm that supports a given social power and the target's experience with the group. Different types of norms can impose additional conditions on Formula (6), which are represented in the norm bias equations below by $NormTypeConditions(n)$. For a position norm that establishes an authority relation, the norm bias is modeled according to Formula (7). Because a position norm ($n$) establishes the right of a member with one role ($r1$) to influence a member with another role ($r2$) in the group, for the corresponding power to be active, the target ($Plays(T, r2)$) and the actor ($Plays(A, r1)$) must be associated with the required roles. $Roles_g$ represents the set of all roles in social group $g$, and $Authority(r_1, r_2)$ represents the authority relation between $r_1$ and $r_2$. Note that a common characteristic of the norm biases associated with all group norms is the salience of the norm. The salience is a probability variable, where a value of 0 represents a low relevance of the norm in the group and a value of 1 indicates that the norm is very relevant in that group. A higher salience of a group norm in a group results in a higher group bias for that norm and, consequently, a stronger overall social power.

$$\begin{aligned} &NormTypeConditions(n) : \exists_{r1,r2} \in Roles_g \wedge Plays(A, r1) \wedge Plays(T, r2) \wedge Authority(r_1, r_2) \\ &\qquad Bias(n) = Salience_n \end{aligned} \tag{7}$$

Similarly, for an action commitment norm[15] that establishes a commitment between two agents with regard to a given behavior, the norm bias is modeled in accordance with Formula (8). Because an action commitment norm ($n$) establishes the right of one agent ($A$) to influence another ($T$) with regard to a given behavior ($C$), for such a power to be active, the corresponding commitment ($ActionCommitment(T, A, C)$) must have been previously established.

$$\begin{aligned} &NormTypeConditions(n) : \exists ActionCommitment(T, A, C) \\ &\qquad Bias(n) = Salience_n \end{aligned} \tag{8}$$

For reciprocity, equity and responsibility norms there is no specific role requirements within the group, but the associated norm bias does not depend solely on the norm salience. A *Reciprocity* norm specifies that when an agent in the group has been benefited by group actions, that agent should also provide a benefit to other members of the group. The norm bias for a reciprocity legitimate social power is modeled in accordance with Formula (9). This social power is affected by the group's history of support (reciprocity bias) toward the target and increases with an increasing percentage of all group actions that have been beneficial to the target agent and were not performed by the target him/herself. This reciprocity bias is a probability variable, where a value of 0 indicates that the agent has never been benefited by group actions and a value of 1 indicates that all group actions have benefited the agent.

$$\begin{aligned} &NormTypeConditions(n) : \varnothing \\ &Bias(n) = Salience_n * \frac{HelpfulActions_{g,T}}{AllActions_g} \end{aligned} \tag{9}$$

An *Equity* norm specifies that when an agent of the group has been negatively affected by group actions, that agent has the right to claim compensation. The norm bias for an equity legitimate social power is modeled in accordance with

[15] Action commitments are interrelated with reward and coercive social powers because in the case of a successful influence interaction involving a social power of the reward or coercive type, a commitment between the participants in the interaction is always generated.

Formula (10). This social power is affected by the group's history of harm (equity bias) toward the actor and increases with the percentage of all group actions that have negatively affected the actor agent. This equity bias is a probability variable, where a value of 0 indicates that the agent has never been negatively affected by group actions and a value of 1 indicates that all group actions have negatively affected the agent.

$$Bias(n) = Salience_n * \frac{HarmfulActions_{g,A}}{AllActions_g} \quad NormTypeConditions(n) : \varnothing \tag{10}$$

A *Responsibility* norm specifies that an agent should help other members of the group who are dependent on that agent. When a social power is based on such a norm, as modeled by Formula (11), the corresponding bias is also proportionally affected by the degree of dependence $Dependence_{A,T,C}$[16] of the actor agent on the target. This bias is also a probability variable, where a value of 0 indicates that the actor does not depend at all on the target's performance of $C$ and a value of 1 indicates that the actor strongly depends on the target's performance of $C$.

$$Bias(n) = Salience_n * Dependence_{A,T,C} \quad NormTypeConditions(n) : \varnothing \tag{11}$$

*Dynamics.* For a legitimate social power, all dynamics occur immediately after the interaction phase regardless of its success. Note that although these effects are not our focus, non-immediate effects may occur in the group-level dynamics (e.g., re-evaluation of group membership) and in the norm-level dynamics (e.g., re-evaluation of norm saliency) [36].

The group's importance and norm salience for the target are defined at authoring time. Additionally, the value of the dutifulness felt by the target depends on the agent's intrinsic values, which are also fixed at authoring time. By contrast, the group norm conformity depends solely on the agent's experience and observation of the conformity of other agents with the established norms in the group. Therefore, this value varies with successes and failures in situations in which such a social power is active. In the case of a successful influence interaction in which a member of the group complies with an established group norm, the norm conformity increases. By contrast, when such an influence interaction is unsuccessful, the perception of the norm conformity decreases because of the failure to comply with the established group norm.

Finally, *Reciprocity*, *Equity* and *Responsibility* norms have additional norm-specific biases that depend solely on the agent's experience with and observation of other group members. Actions performed by members of the group that are beneficial to an agent cause the reciprocity bias of that agent (target) to increase. Similarly, actions performed by members of the group that are harmful to an agent cause the equity bias of that agent (actor) to increase. All other actions performed in the context of the group indirectly result in a decrease of both of these norm-specific biases.

*Examples of critical incidents.* Many social groups, both institutional and informal, can be found throughout human societies. However, given the typical focus on institutional perspectives, although this model can express both, the critical incidents we will explore relate to the less emphasized but also very important class of informal groups. Imagine a social group of the kind that typically emerges in a family. In such a context, it is common for there to be position-related norms, such as norms establishing that the younger members of the family should obey the older members. Thus, if a child feels strongly duty-bound to comply with this group norm, then the legitimate social power arising from this norm will be very strong. For example, this occurs in a situation in which the child is going out with some friends and the parent asks him to be home before midnight. If, on the contrary, the child feels little dutifulness, then he might easily refuse to obey the parent's instruction and thus fail to comply with the family group position norm when in the presence of other social or personal forces.

In the same setting, if the members of the family group frequently fail to comply with family norms, then the legitimate social power of the family group position norm is strongly weakened. For instance, the child might claim that he will not obey the parent's instruction to be home by midnight because the other family members frequently fail to comply with other informal norms existing in the family (e.g., requirements to provide for the family). By contrast, if everyone in the family group always complies with group norms, then this legitimate social power component exerts a positive force toward enhancing the strength of the legitimate social power.

Similar situations can be easily imagined for specific action commitment norms. For instance, consider a family in which it has been established that a specific child has the task of setting the table for dinner every day. In all such situations, the social power can strongly vary with the importance of the group to the target and the salience of the norm within the group. If, for example, the target considers his/her association with the family to be very weak (e.g., mostly disregards family interactions), then this social power is strongly diminished and may even be nullified. Similarly, if a norm has little salience in the group, then the social power resulting from it is diminished.

[16] We do not specify how dependence should be quantified because this as a complex subject in itself and strongly depends on context. However, in general, the degree of dependence should be inversely proportional to the number of alternatives available to the actor regarding the target's performance of $C$ [28].

Reciprocity and equity norms may also be found in a family group. As an example of a reciprocity norm, imagine a norm establishing that family members should help each other when possible. Hence, a member of the family who is frequently helped by other members of the family experiences a stronger legitimate reciprocity social power toward helping other members of the family group. Conversely, if no one in the family ever helps a particular family member, then that member will feel less (or not at all) compelled, out of reciprocity, to perform an action for another member of the family. Similarly, an equity norm may also exist in a family establishing that its members should equally share personal disadvantages among the group. If, for instance, the family is on vacation in a foreign city and have visited many attractions, but none that is the preference of a particular family member, then that member has a strong legitimate equity social power over the other members of the group to influence them to visit an attraction of his/her choice.

A family group scenario also offers examples of critical incidents involving responsibility norms. For example, imagine a situation in which a child who is going out with friends lives far from the place where he has arranged to meet with them and public transportation is scarce. When the child asks a parent to drive him to the meeting place, because he has few or no alternatives, the parent experiences a strong legitimate responsibility social power influencing him/her to help the child. This occurs because the family upholds a norm that family members should help each other when they are in need (i.e., when they have few or no viable alternatives).

#### 4.3.4. Referent

In an influence situation, a referent social power exists when a target ($T$) identifies him/herself with an actor ($A$) based on some personal relation ($Liking_{T,A}$) or status recognition for the actor in a given status category that compels the target, based on his/her personal motivations, to comply with the actor. Note that although status has interdependencies with other types of social powers (e.g., legitimate), we model it as a referent social power because of its operationalization mechanism and its conceptual link to identification factors [36].

A referent social power as modeled in this work can be of one of two types: liking or status. A referent liking social power that an actor exerts over a target that influences the latter to perform an action is a function of the strength of the liking that the target holds for the actor ($Liking_{T,A}$) and the target's personal affiliative motivation ($AffiliativeM_T$). This description is succinctly modeled by Formula (12).

$$\begin{aligned} PB_{T,ReferentLike} &= \{AffiliativeM_T\} \\ Be_{T,ReferentLike} &= \{Liking_{T,A}\} \\ f_{ReferentLike}(PB_{T,ReferentLike}, Be_{T,ReferentLike}) &= Liking_{T,A} * AffiliativeM_T \end{aligned} \tag{12}$$

A referent liking social power is therefore directly proportional to the strength of the liking relation held by the target agent and to its affiliative motivation. The strength of a liking relation is a positive/negative variable, where a negative value represents a dislike situation, a positive one represents a like situation, and a value of 0 represents an indifferent relation regarding liking. The affiliative motivation of the agent is a probability variable, where a value of 0 indicates that the agent has no such motivation and a value of 1 indicates that the agent has a very strong affiliative motivation.

The other type of referent social power is based on social status, which is defined as a representation of an individual's social prominence and respect [1]. The referent status social power that an actor exerts over a target that influences the latter to perform an action is a function of the actor's status ($Status_{T,A,cat}$) as recognized by the target, the target's personal status motivation ($StatusM_T$) and the target's preference for the specific category of status held by the actor ($Importance_{cat}$). This description is succinctly modeled by Formula (13). Note that *StatusCategories* is the set of all status categories that the target values.

$$\begin{aligned} &\exists_{cat} \in StatusCategories \\ PB_{T,ReferentStatus} &= \{StatusM_T\} \\ Be_{T,ReferentStatus} &= \{Status_{T,A,cat}, Importance_{cat}\} \\ f_{ReferentStatus}(PB_{T,ReferentStatus}, Be_{T,ReferentStatus}) &= Status_{T,A,cat} * StatusM_T * Importance_{cat} \end{aligned} \tag{13}$$

A referent status social power is therefore directly proportional to the level of status ascribed by the target to the actor, the target's status motivation and the target's preference for the relevant status category (e.g., popularity in a group of friends). The recognized level of status is a positive/negative variable, where a negative value represents a bad status situation (e.g., unpopularity), a positive one represents a good status situation (e.g., popularity), and a value of 0 represents an indifferent status standing. The status motivation of the agent is a probability variable, where a value of 0 indicates that the agent does not value status relations and a value of 1 indicates that the agent strongly values status relations. Finally, the category preference bias is another positive/negative variable, where negative values indicate that the agent negatively regards that type of status, positive ones indicate that the agent positively regards that type of status, and a value of 0 indicates that the agent is indifferent to that type of status.

*Dynamics.* For a referent liking/status social power, all dynamics occur immediately after the interaction phase regardless of its success.

Regarding a referent liking social power, the value of the liking relation held by the target agent depends solely on the target's interaction experience with the actor agent. Nonetheless, such relations are generally affected in both valence and magnitude by the target's evaluations of others' actions with respect to the target's goals, identity and social relations [10, 22]. Additionally, in accordance with French and Raven's model [37], a liking relation is also strengthened by reward actions (depending on reward execution) and weakened by coercive interactions (regardless of coercion execution). Meanwhile, the affiliative motivation magnitude of the target agent depends on that agent's intrinsic value, which is fixed at authoring time because it is typically considered to be a stable characteristic.

Regarding a referent status social power, the value of the actor's status as recognized by the target agent is most strongly dependent on the target's experience of interaction with the actor agent and but is also affected by indirect exchanges regarding the actor with other agents. However, as a general indication for the authoring of such social relations in an agent system, status values are frequently affected by the performance of specific actions or the attainment of specific goals [10]. The status motivation magnitude of the target agent depends on that agent's intrinsic value, which is fixed at authoring time because it is typically considered to be a stable characteristic. Finally, the category preference magnitude is also an intrinsic value for an agent that is set at authoring time. Although certain types of category preferences [45] are not continuously changing, they are not as stable as other characteristics, such as status motivation [57].

*Examples of critical incidents.* Imagine a situation in which Nick, a friend of Mary, asks her to help him to study for a test. Nick's friendship for Mary is represented by a high liking relation between the two. This relation strengthens the referent liking social power imposed on Mary that influences her to accept the request to help Nick study. If, on the other hand, Mary disliked Nick, she would instead be compelled not to help because of that dislike. Regarding the affiliative motivation, if Mary has a high need for affiliation, this means that she strongly values her liking relations and therefore that the referent liking social power will be strengthened in either case, whether positive (if, for example, Mary likes Nick) or negative (if Mary dislikes Nick). In the case of a null affiliation motivation, Mary does not value her affiliation relations at all and, therefore, no social power will be derived from her liking relation with Nick.

Now imagine another situation in which Jane, a popular member of a school class, asks Carl, who is an unpopular classmate, to lend her some money. The status category that is relevant to this example is popularity. If Carl recognizes Jane's popularity status and has a high regard for popularity status, then the status-based referent social power that is active will be very effective. However, if Carl has little regard for status in general or for popularity status specifically, this will result in a weak referent status social power influencing him toward lending Jane money. Alternatively, if Carl does not recognize Jane's high popularity status, then even if he has a high regard for status and values popularity very much, the status referent social power influencing him toward lending the money will again be very low.

#### 4.3.5. Expert

In an influence situation, an expert social power exists when a target ($T$) maintains a trust relationship[17] with an actor ($A$) and simultaneously recognizes that the actor has a superior skill ($s$) associated with the action under influence.

An expert social power that an actor exerts over a target that influences the latter to perform an action is a function of the importance of the skill associated with that action ($Importance_{s,C}$), the trust that the target agent holds for the actor agent ($Trust_{T,A}$) and the target's perception of the skill difference between the actor and the target ($LevelD_{A,T,s}$) regarding the skill associated with the action. This description is modeled by Formula (14). Note that *Skills* represents the set of all known skills.

$$\begin{aligned}
\exists_s &\in Skills \\
PB_{T,Expert} &= \{\} \\
Be_{T,Expert} &= \{Importance_{s,C}, Level_{A,s}, Level_{T,s}, Trust_{T,A}, MaxSkillLevel_s\} \\
LevelD_{A,T,s} &= \frac{Level_{A,s} - Level_{T,s}}{MaxSkillLevel_s} \\
ExpertB &= \begin{cases} Trust_{T,A} * LevelD_{A,T,s}, & \text{if } Trust_{T,A} \neq 0 \wedge LevelD_{A,T,s} > 0 \\ 0, & \text{otherwise} \end{cases} \\
f_{Expert}(PB_{T,Expert}, Be_{T,Expert}) &= Importance_{s,C} * ExpertB
\end{aligned} \tag{14}$$

The strength of an expert social power is therefore directly proportional to the importance of the associated skill, the magnitude of the target's trust relationship with the actor and the recognized skill difference between the actor and the target. The skill importance specifies the relevance of the current skill to the action under influence and is a positive variable, where values close to 0 represent a skill that is very weakly associated with $C$ and higher values indicate that the skill is more strongly associated with $C$. The trust that the target holds for the actor depends on the target's experience with the actor and is a positive/negative variable, where positive values indicate that the target trusts the actor, negative values indicate that the target distrusts the actor, and a value of 0 indicates that there is no trust relationship. The perceived skill

[17] In this definition, we use a conceptualization of trust from a reliance perspective [31].

level also depends on the target's experience with the actor and is a positive variable, where values close to 0 indicate that the actor agent is perceived as unskillful with regard to the associated skill and a larger magnitude of the variable indicates a greater level of skill on the part of the actor agent. The quantity $MaxSkillLevel_s$ identifies the maximum skill level known to the target agent.

*Dynamics.* For an expert social power, all dynamics occur immediately after the interaction phase regardless of its success.

The skill importance is a value that is set at authoring time because it represents a prototypical and stable perception. A trust relationship depends on the experience of the target agent with the actor agent with respect to actions that increase/decrease trust according to specific dynamics [31]. Regarding the perceived skill, changes occur when actions associated with learning (e.g., when someone reads a book about programming) or demonstrations of skill (e.g., when someone programs something very difficult and others recognize that person's skill) are performed [10].

*Examples of critical incidents.* A typical situation in which an expert social power may be observed is when a physician instructs a patient to follow a given medical prescription. In this situation, Bart, the patient, recognizes that medical skills are important for prescribing an appropriate medicine. Even if, for instance, Bart knows that John (the physician) has excellent musical skills, this knowledge exerts no expert social power in this interaction because there is no association between musical skills and effective medical prescriptions.

If Bart maintains a positive trust relationship with John and Bart recognizes that John is a very good doctor (and thus has very good medical skills), then John's expert social power is increased. By contrast, if Bart knows that John has, for instance, not yet graduated with his medical degree, then he might be regarded as an unskillful doctor, causing the expert social power to be strongly diminished.

Now, imagine a situation in which Bart has been warned that John tends to favor certain pharmaceutical companies in his prescriptions, regardless of the best course of treatment (e.g., treatments that minimize secondary effects). Although John is undoubtedly more skillful than Bart in medicine, the fact that he is using those skills not for Bart's benefit but for his own benefit makes John unreliable and, thus, a source of negative expert social power for Bart. Note that in this case, it is not the main underlying factor of the expert social power (skill) that is responsible for the negative change in social power but rather the trust factor, which directly impacts the main factor.

### 4.4. Putting it all together: a comprehensive example

Consider a situation in which a boy is going out with his friends, but before leaving home, he is instructed by his father to be home before midnight. At a certain point after leaving home and before midnight, he will be faced with the decision to either do as told by his father or defy his father's wishes. What are the forces in play affecting the boy's decision? How do we apply our model?

#### 4.4.1. Initial situation

The agents in this scenario are $A = father$ and $T = boy$, and the action upon which the *boy* will have to decide is $C = ReturnOnTime$. Although there can be many established social norms in a *Family* group, for this example, we will consider only one (informal) position social norm that is frequently adopted in families: children should always obey their parents (*ObeyParents*). Based on this, we have the *father* assuming the parent role and the *boy* assuming the role of the child. Regarding liking relations, we assume that the *boy* likes his *father*. Concerning actions, we consider the father's capability to *WithdrawAllowance*, *GroundBoy* and *AllowReturnLate*. The *boy*'s action that is relevant to this example is his capability to *ReturnOnTime*. Finally, the decision that is the focus of this example is that made by the boy, in which he must choose to either $Do(boy, ReturnOnTime)$ or $\neg Do(boy, ReturnOnTime)$.

#### 4.4.2. Modeling the boy's decision

Following the introduced social power decision mechanism (see Section 4.2) and social power dynamics (see Section 4.3), the *boy* begins by identifying the social powers that are active in his current situation. As a result, the boy identifies the following potential social-power forces:

1. $PotentialForce_{coe1} = f_{Coercive}(\{AchievementM_{boy}\}, \{Coercing_{boy,father,WithdrawAllowance}\})$
2. $PotentialForce_{coe2} = f_{Coercive}(\{AchievementM_{boy}\}, \{Coercing_{boy,father,GroundBoy}\})$
3. $PotentialForce_{rew1} = f_{Reward}(\{AchievementM_{boy}\}, \{Rewarding_{boy,father,AllowReturnLate}\})$
4. $PotentialForce_{leg1} = f_{Legitimate}(\{DutifulnessM_{boy}\}, \{Importance_{Family}, Salience_{ObeyParents}, Conf_{boy,Family}\})$
5. $PotentialForce_{ref1} = f_{ReferentLike}(\{AffiliativeM_{boy}\}, \{Liking_{boy,father}\})$

The coercive powers represent the *father*'s ability to punish the *boy* by withdrawing his allowance (1) or grounding him (2) from future opportunities to go out with his friends. Concerning rewards, the *father* can allow the *boy* to stay out longer (3) the next time if he behaves properly this time. The legitimate power (4) represents the *father*'s right to influence the *boy* regarding *ReturnOnTime* arising from the fact that the *boy* plays the role of the child and the *father* plays that of the parent. The referent power (5) is based on the son–father relationship between the two agents. No expert power exerts any force because there is no skill that is relevant to this situation.

**Table 2**
Example values for the case of a well-behaved child.

| Characteristic | Value | Characteristic | Value |
|---|---|---|---|
| $AchievementM_{boy}$ | 0.6 | $Value_{GroundBoy}$ | −200 |
| $DutifulnessM_{boy}$ | 0.7 | $Coercing_{boy,father,GroundBoy}$ | 0.3 |
| $AffiliativeM_{boy}$ | 0.8 | $Importance_{Family}$ | 100 |
| $Value_{AllowReturnLate}$ | 50 | $Salience_{ObeyParents}$ | 0.9 |
| $Rewarding_{boy,father,AllowReturnLate}$ | 0.4 | $Conf_{boy,Family}$ | 0.8 |
| $Value_{WithdrawAllowance}$ | −150 | $Liking_{boy,father}$ | 100 |
| $Coercing_{boy,father,WithdrawAllowance}$ | 0.1 | $ValueForce(boy, ReturnOnTime)$ | −100 |

The *boy*'s decision between $Do(boy, ReturnOnTime)$ and $\neg Do(boy, ReturnOnTime)$ thus depends only on the values of the forces that are active in this situation. From the perspective of what occurs in real life, both cases are believable given the appropriate personal characteristics, relationships and beliefs. We will present both cases accordingly.

#### 4.4.3. Case 1: a well-behaved child

If we assume a well-behaved son, then all of the social-power forces work toward $C$, and although the value to the *boy* of the action of returning on time offers resistance to this decision, he will return home before midnight. To illustrate this case, consider the values given in Table 2. Regarding personal characteristics, we assume that the *boy* is moderately achievement motivated, strongly dutiful and strongly affiliation motivated. Additionally, the *boy* considers the father to have a low tendency to withdraw his allowance and moderate tendencies to ground him and allow him to return late. He positively values the *father*'s capability to allow him to return later at another time. By contrast, he very negatively values the *father*'s capabilities to withdraw his allowance and ground him. Concerning his group affinity, we assume that the *boy* values his *Family* group and that within the family, it is very important to uphold the existing position social norm. In this context, the *boy* perceives the *Family* group to have a high norm conformity. The liking relationship that the *boy* has with his *father* is strong. Finally, the *boy*'s resistance to returning on time is also strong.

Note that the values assigned to each characteristic are mapped in accordance with the described situation and with consideration for the associated variable type for each assignment. For instance, "low tendency to withdraw allowance" is mapped to a probability-like variable with a value of 0.1, whereas "moderate tendencies to ground and to allow late return" are mapped to probability-like variables with values of 0.3 and 0.4, respectively. In this way, we maintain the meaning conveyed in the example but express it in a quantitative form that can be used in the model. Other types of parameters that are used are characterized as either positive (e.g., the importance of the family social group) or positive/negative (e.g., the boy's value evaluation of having his allowance withdrawn). These types of parameters are not restricted by the model in terms of upper/lower bounds, although these bounds should be considered when authoring the scenario. The most important precaution to consider is that the numerical values must maintain relative values between them that reflect the intended semantic differences depicted in the scenario. For example, considering the value of −150 assigned to the boy's evaluation of his allowance being withdrawn by his father, which is negative because of the loss of resources it can cause; however, the value associated with being grounded, −200, is even more negative because this action would very strongly restrict the boy. However, we can easily imagine that neither of these situations is valued nearly as negatively as becoming severely injured (e.g., −900).

Based on Table 2, we can calculate the actual values of the identified potential social-power forces:

$$PotentialForce_{rew1} = 50 * 0.6 * 0.4 = 12$$

$$PotentialForce_{coe1} = -(-150) * 0.6 * 0.1 = 9$$

$$PotentialForce_{coe2} = -(-200) * 0.6 * 0.3 = 36$$

$$PotentialForce_{leg1} = 100 * (0.7 + 0.8) * 0.9 = 135$$

$$PotentialForce_{ref1} = 100 * 0.8 = 80$$

Now, according to definition (2), it is possible to calculate the overall social-power force:

$$InfluenceForce(boy, father, ReturnOnTime) =$$

$$= PotentialForce_{rew1} + PotentialForce_{coe1} + PotentialForce_{coe2} + PotentialForce_{leg1} + PotentialForce_{ref1} =$$

$$= 12 + 9 + 36 + 135 + 80$$

$$= 272$$

Now that we have both the social-power force and the value force, we can calculate the resultant force and determine the decision of the *boy* agent according to (3). For this example, we assume a balance between the value and influence forces of $\alpha = 0.5$. Therefore, the resultant force value is as follows:

$$ResultantForce(boy, father, ReturnOnTime) = 0.5 * -100 + (1 - 0.5) * 272 = 86$$

Based on the value of the resultant force and the condition $ResultantForce(boy, father, ReturnOnTime) > 0$ from definition (3), the *boy* agent decides on $Do(boy, ReturnOnTime)$, meaning that he will return home on time, before midnight.

#### 4.4.4. Case 2: disobedience in a well-behaved child

Now imagine that the *boy* is having more fun than he has ever had before. In this case, the situation might arise in which a well-behaved son nevertheless disobeys his father's command because the value force surpasses the social-power-based influence force exerted by his father. This means that the evaluation of the situation changes drastically. To illustrate this scenario, suppose that the value of $ValueForce(T, C)$ in Table 2 changes from $-100$ to $-300$, while all other values remain the same. Because the underlying factors associated with the social power calculation do not change, the resultant force including the impact of this change is as follows:

$$ResultantForce(boy, father, ReturnOnTime) = 0.5 * -300 + (1 - 0.5) * 272 = -14$$

Based on the value of the resultant force and the condition $ResultantForce(boy, father, ReturnOnTime) \leq 0$ from definition (3), the *boy* agent decides on $\neg Do(boy, ReturnOnTime)$, meaning that he will not return home on time and will stay out late.

#### 4.4.5. Other complex cases

An additional interesting case is that of a rebellious son, in which the legitimate component of the social-power force would exert a negative influence because the command issued by the father actually exerts a power that works against *ReturnOnTime*. In this case, the rebellious son is much more likely to stay out with his friends $\neg Do(boy, ReturnOnTime)$ even in the first case. Note that this does not mean that he "does not like" his father; a son might like his father and simultaneously exhibit rebellious behavior. Our model enables the simulation of such complex situations.

## 5. Integrating SAPIENT with typical agent processes

To develop agent systems implementing the model described above, we must understand the links between typical agent processes and the introduced social power model. From a computational perspective, our main inspiration for the creation of an agent framework to support the SAPIENT model is the BDI architecture [8,65]. Based on this architecture, our starting point is a typical agent with corresponding sets of Beliefs (knowledge about itself, others and the environment), Desires (goals to pursue) and Intentions (actions or plans to which the agent is committed). The SAPIENT model is then integrated into the deliberative process of the agent, from the generation of options to the actual selection of an option and belief updates. In this section, we present a cognitive agent architecture, focusing on the links between these typical BDI agent structures and social-power-related concepts.

To create an agent architecture that can endow agents with social power awareness and the ability to generate behaviors based on plans while incorporating social power strategies, our main inspiration is Raven's Power/Interaction Model [66]. The advantage of using this model is that it was specifically created to explore the links between the bases of social power introduced by French and Raven [37] (and used in the SAPIENT model) and their operationalization in influence interactions. In a social power interaction, there are two main interaction perspectives, each requiring a specific cognitive process. One is the perspective of the influencing or actor agent (see Fig. 5), and the other is the perspective of the influenced or target agent (see Fig. 6). Although the descriptions of the SAPIENT model given above were fundamentally focused on the target of the interaction, in this section, we describe agents that can be either targets or actors.

### 5.1. Perspective of the actor agent

The actor agent's cognitive process (see Fig. 5) begins with the evaluation of its motivations (needs or goals). When an agent has a motivation to influence another (e.g., needs another to perform some action), he/she evaluates the bases of social power that are relevant for influencing that other agent. Based on this initial analysis, the agent decides whether a strategic interaction is required to emphasize his/her social power and improve the chances of a successful influence interaction.[18] If the agent decides to use a social power strategy, he/she assesses which strategies he/she can employ given the available bases of social power. During the "Choice of Influence Mode" cognitive stage, the agent then chooses his/her preferred power strategy (if any) and attempts to exert his/her social power by communicating its influence to the target agent. When refraining from using a social power strategy the communication of the influence does not in any way emphasize any social power. After a reply from the interaction target, the actor assesses the effects resulting from the influence interaction (e.g.,

[18] Power strategies are related to "Preparatory Devices" in the Power/Interaction Model [66].

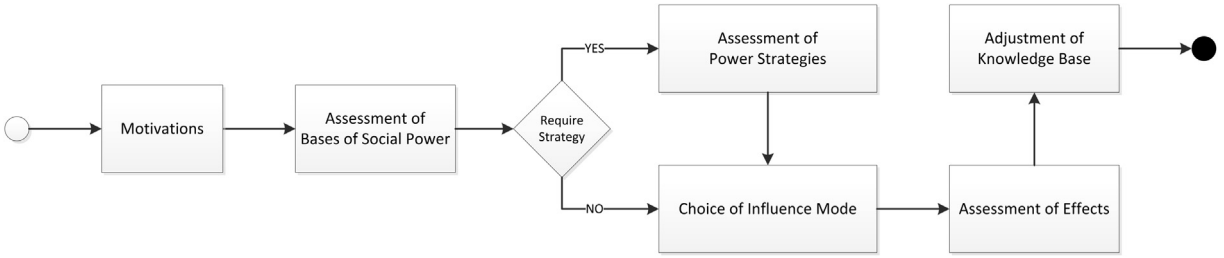


**Fig. 5.** The actor agent's cognitive process in a social power interaction episode.

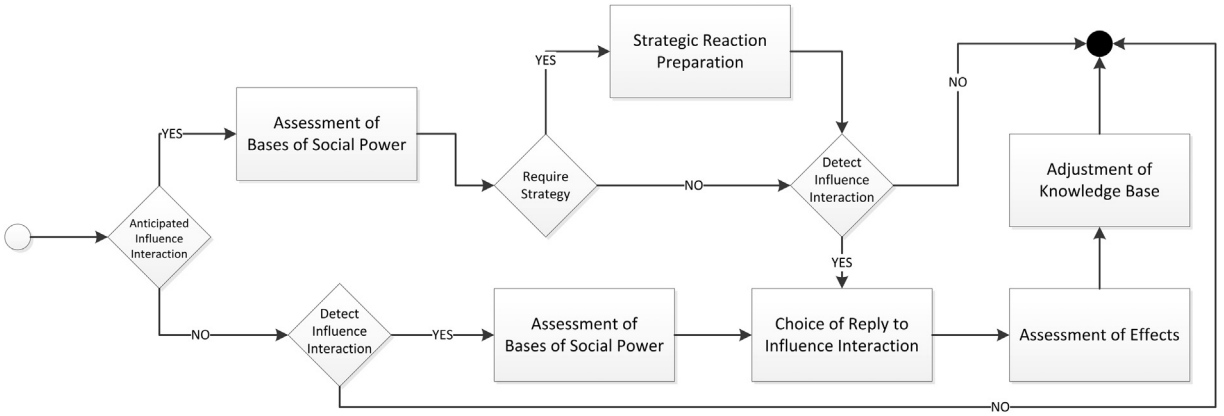


**Fig. 6.** The target agent's cognitive process in a social power interaction episode.

an increase or decrease in interpersonal attraction [37]) and adjusts the social power structures in his/her Knowledge Base accordingly.

In a social-power-based interaction, one agent can influence another in two different ways. First, the agent might simply ask another agent to do something. The second possibility is to ask for that same thing while simultaneously emphasizing a relevant social power base, thereby increasing the magnitude of the exerted power. These modality considerations are called social power strategies and are used to emphasize one or several bases of social power. Power strategies are diverse [35] and involve various types of considerations (e.g., the highlighting of a capability for a possible coercive action). Finally, it is important to remark that the actor can also passively influence the target without being aware of it and, therefore, without engaging in the cognitive process described above. This situation occurs when the actor is not aware of the influence that it is exerting on the target agent. For instance, imagine a police officer standing on a sidewalk during a work break, thereby inadvertently dissuading illegal parking.

### *5.2. Perspective of the target agent*

The target agent's cognitive process (see Fig. 6) begins with an analysis of the agent's motivations and past interactions with other agents. If no influence interaction is anticipated but one occurs nonetheless, then the agent begins by assessing the bases of social power identified as relevant to that interaction. Based on that assessment, the agent decides to either accept or reject the influence interaction. Next, similarly to the actor agent, the target will then assess the effect resulting from the interaction and adjust his/her Knowledge Base accordingly. In the case in which the agent anticipated the influence interaction, he/she plans and prepares a reaction to the anticipated influence interaction. The first step is the assessment of the relevant social power bases for the interaction. Next, the agent decides whether he/she requires[19] a strategy to resist the anticipated influence interaction. At this stage, the agent has planned a reaction to the anticipated influence interaction, and if and when that interaction actually occurs, the agent reacts to it with the planned reaction. Following the reaction, the agent assesses the effects and updates his/her Knowledge Base accordingly.

### *5.3. Cognitive agent architecture*

To operationalize the presented cognitive processes, we developed an agent architecture for agents that are socially intelligent with regard to social power. The developed architecture, presented in Fig. 7, highlights the relations of the typical

[19] There are situations in which resistance is not desired, for instance, if the agent would benefit from accepting the influence interaction.

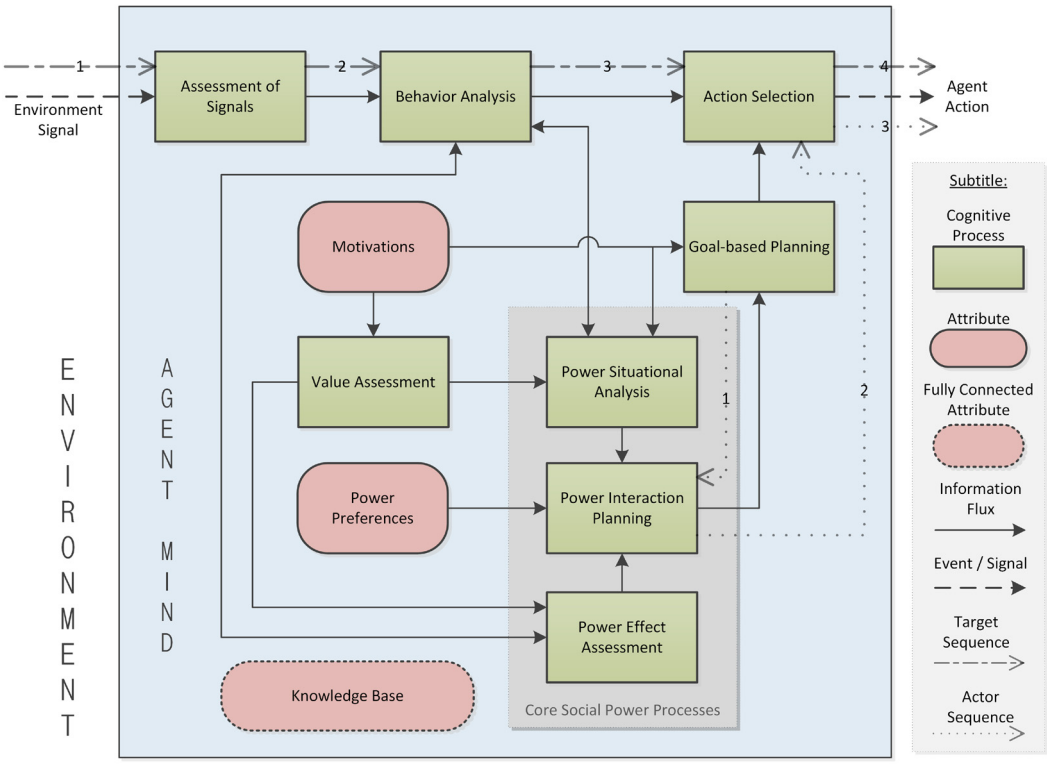


**Fig. 7.** Agent architecture depicting the main fluxes of information among the modeled cognitive processes. Although the connections are not directly depicted for clarity of presentation, the Knowledge Base is represented as a fully connected attribute because it is connected to all other processes.

agent processes with three core social power processes that are identified as fundamental for agents participating in social power interaction episodes. Note the main sequence of cognitive processes for both actor and target agents (without anticipation), which maps the previously presented fluxograms to the developed architecture. The Knowledge Base is graphically represented as a fully connected attribute because it is a fundamental component upon which all other processes depend, and as such, although it is connected to all other processes, these connections are not explicitly represented for clarity of presentation.

Considering the underlying inspiration provided by the typical BDI architecture [8,65], in the following sections, several components are discussed in further detail, with a focus on their relations to social power processes. The key implications of extending such an architecture to include social power processes are related to the structure of the agent's beliefs and individual characteristics and how the agent's deliberative process is affected. As such, several processes, although important for a fully operational agent, will not be discussed in detail because they require no significant changes regarding their function. This is the case for the following processes: Assessment of Signals (sensing of signals from the environment and other agents), Goal-Based Planning (typical planning process for an agent's action) and Action Selection (coordination of agent actuation on the environment).

### *5.3.1. Beliefs, motivations and preferences*

Some of the key changes that social power processes entail in a typical agent architecture are related to the agent's belief structure and the attributes that characterize the agent's individual behavior. These are used in the deliberative processes to support social-power-related analysis and impact the selection among different options. The components focused on in this section are the Knowledge Base, Motivations and Power Preferences.

An agent's Knowledge Base is, effectively, the agent's internal source of information, which is used and updated by all other agent processes. The Knowledge Base structure, presented in Fig. 8, highlights the different types of knowledge that have an impact on social-power-related agent processes. The "General Beliefs" attribute includes all knowledge that is necessary to make an agent aware of its environment (e.g., the positions of objects). The "Models of Others" attribute represents an agent's beliefs regarding other agents, i.e., the agent models other agents as similar cognitive agents. A model of another agent is built using information collected from direct interactions with that agent and from observations of interactions between the modeled agent and others. Additionally, in the absence of information regarding another agent's motivation characteristics, these are projected from the agent's own personal characteristics (a form of human projection bias [61]). By modeling other agents in this manner, the agent can reason about situations from others' perspectives and integrate that information into its own analysis and decision processes (using the Behavior Analysis process). The "Goals"

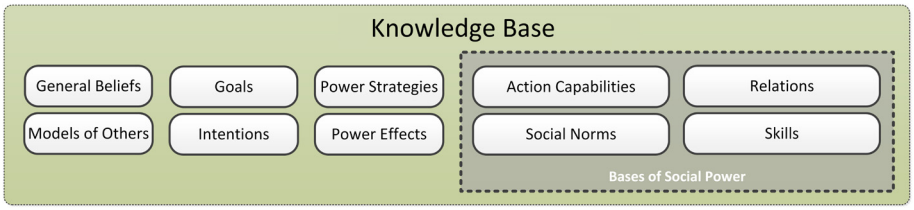


**Fig. 8.** Knowledge Base structure for a social-power-intelligent agent.

attribute includes the agent's mental representations of desired world states. These range from non-social goals (e.g., move a chair) to social goals (e.g., harm another agent). The "Intentions" attribute represents the actions that the agent has committed to perform, e.g., actions to complete a particular plan. The "Power Strategies" attribute represents the agent's known strategies for using various bases of social power. These have specific constraints, preference biases and modes of usage [34,36]. The "Power Effects" attribute represents the agent's knowledge about the effects that the usage of various social power bases and strategies can have on its beliefs (e.g., an increase/decrease in a liking relation).

Next, in our social power model, a group of beliefs that underlie the bases of social power is defined. The "Action Capabilities" attribute represents the actions that an agent can perform (e.g., provide a resource) and that underlie coercive and reward bases of social power. The "Social Norms" attribute represents all types of internalized values that support legitimate bases of social power, modeled as either formal or informal social norms [12]. The "Relations" attribute registers the different types of social relations (i.e., liking and status) that an agent maintains with others and serves as the basis of referent social powers. Finally, the "Skills" attribute represents the skillfulness of the agent, which is fundamental for determining the expert social powers associated with actions requiring specific skills to be performed effectively. For instance, giving an object to someone is a typical action that does not require any expertise; by contrast, the programming of a computer application can only be performed by a person with computer programming skills.

Another component that impacts the agent's perceptions of social power is the agent's Motivations. The agent's motivational structure is a combination of social-power-relevant factors and scenario-relevant factors. Scenario-specific motivation factors depend on the implemented setting and must be specified when creating the scenario. For the social-power-relevant factors, the Power/Interaction Model [66] provides the general motivational framework by identifying social-power-specific motivations such as goals and the satisfaction of internal needs. However, this model does not clarify which internal needs can be used to characterize an agent.

For this purpose, we use McClelland's Human Motivation theory [57] because of its clear intersection with our conceptualization of social power. Social-power-specific motivations have three core dimensions. The first is achievement and is related to the need to accomplish goals, which directly impacts reward and coercive social powers as well as the value assessment process of the agent. The second is affiliation and is related to the need to nurture and uphold social relations, which is directly related to referent liking social powers. The third motivation dimension is power and is related to the need for status and recognition; therefore, this dimension is used to model the personal status motivation of the agent relevant to referent status social powers. Even with the invocation of McClelland's Human Motivation theory, support for one intrinsic agent motivation is still lacking: the dutifulness of the agent. To address this lack, we have found that we can model dutifulness as a combination of the personality traits of agreeableness and conscientiousness [36] from the five-factor model [19]. All these factors map the personal biases described in the developed social power model.

The final attribute that is intimately related to the individual behavior of an agent is the agent's Power Preferences. Although in general, individual preferences influence most actions to some extent, in considering this attribute, we model only those preferences that are related to the agent's utilization (from the actor's perspective) of social power (e.g., preferences regarding the use of either coercive or referent social power bases). Again, social psychology offers valuable insights into the dimensions that can be used to model these individual preferences. One of the factors that has been specifically correlated with the utilization of different bases of social power is personality [34,36]. In modeling this attribute, we map an agent's personality characteristics to biases toward using strategies relying on particular social power bases in their social interactions according to the findings of Karkoulian et al. [44]. Their study was based on the five-factor model [19] and addressed French and Raven's [37] bases of social power, and as such, their findings can be directly mapped to the preference mechanism in the developed model.

### *5.3.2. Core social power processes*

To create an agent that is capable of sensing, deliberating and taking action based on social power, we must understand how social power is manipulated. This occurs at three different levels, each of which is represented in the SAPIENT architecture as a core social power process. This section focuses on examining these core processes and their links with other agent components.

The first core social power process is the Power Situational Analysis process, the primary purpose of which is to identify and quantify the social-power forces that are relevant to a given (possible or anticipated) influence interaction. One fundamental force directing the agent's behavior is the value of the action underlying the influence interaction. To determine this value and use it in this assessment, the agent uses the "Value Assessment" process, which conceptualizes the agent's

capacity to assess the value of specific actions associated with a social-power-based interaction according to its beliefs, motivations and goals. At a basic level, this assessment measures the value of the action for the agent based on concepts similar to those of utility in the sense that it evaluates the usefulness of performing the action. However, the focus of the Power Situational Analysis process is the potential social-power forces that can either oppose or reinforce the value of the action. To assess these forces, the agent must detect the social powers modeled in Section 4 and identify which social powers are active in the current interaction context. For instance, for a reward social power to be active, there must be a rewarding tendency associated with the action underlying the influence attempt. The quantification of social powers is thus a function of all these factors, in accordance with the mechanism introduced in Section 4.2. The final output of this process is the set of all social powers identified as relevant to a given interaction and the resultant influence force that arises from them with regard to an influence interaction. For the pseudocode of this cognitive process, see Algorithm 1 in Appendix A.

The second core social power process is the Power Effect Assessment process; its main purpose is to identify the effects (or outcomes) of a social power interaction. Some inspiration for the modeling of these effects comes from French and Raven [37]. In their presentation of the bases of social power, they discuss the effect of the increase in referent social power that occurs when a reward social power is used and its corresponding decrease when coercion is used. Interestingly, these dynamics can be altered when these social powers are used in combination with legitimate social power. Consider a situation in which a given base for legitimate social power (e.g., a contract or a norm) between two agents may define the legitimacy of a given reward or coercion. Rewards or punishments with different legitimacies will have different effects on the target agent. In legitimate coercion (e.g., the issuance of a fine for running a red light), the negative effect of using coercion is attenuated, whereas in the case of a legitimate reward (e.g., the provision of a bonus for increased productivity), there is no difference.

Although French and Raven associate certain social power effects with specific bases of social power, a general social power dynamics mechanism relating social power interactions with their effects is not discussed. To address this lack of a generic mechanism, which inherently strongly limits the modeling of dynamics to scenario-specific authored effects, we introduce a claim/conferral mechanism inspired by Kemper's Status-Power Theory [46]. This mechanism provides a generic mechanism for social power effects that can be applied to the diverse constituents of Raven's social power conceptualization. For instance, with regard to an expert social power, if a given actor uses an expertise strategy (e.g., stating "You should do this because I have more expertise on this topic"), that actor is making a claim of his higher expert social power. If the target accepts this claim, then the target is conferring a higher expert social power (compared with the perception of that power before the interaction) to the actor, and the claim becomes permanent. If the claim is rejected, this means that the target confers less social power than the actor expected. Similarly, even in the context of a non-strategic interaction (a simple request to perform a given task without emphasis on any factor related to a social power base), if the target recognizes the higher expert social power of the actor (confers a higher expert social power to the actor), then the actor's claim is simply met. In the case of a refusal, the target confers a lower expertise to the actor. This reasoning can be applied to several of the social power bases introduced by French and Raven. The Power Effect Assessment process identifies the changes in social-power-related beliefs that result from a given interaction. For the pseudocode of this cognitive process, see Algorithm 7 in Appendix A.

The last core social power process is the Power Interaction Planning process, and its purpose is to perform online planning for social-power-based interactions. It enables the agent to reason about possible influence situations and select its best option for influencing other agents by integrating all of its knowledge about social powers and their effects as well as its own social power utilization preferences.

The basic information needed for planning an interaction is obtained from the "Power Situational Analysis" process through an assessment of the potential social-power forces that are active in a given context (using the mental model of the target). This information is then used in the "Power Effect Assessment" process to identify the anticipated changes in social-power-related beliefs and analyze their values for the agent. For example, the use of a threat, which emphasizes a coercive base of social power, will unquestionably cause its target to dislike the agent using it, thereby diminishing that agent's referent social power over the target. Once the potential social-power forces and anticipated effects have been identified, the agent must still compare and choose among the different interaction possibilities. It is at this stage of planning that the agent's "Power Preferences" attribute biases the agent's choice of strategy by increasing the values of certain interaction possibilities in accordance with personality-related biases [44]. Once these biases are applied to all interaction possibilities, the agent has a range of options from which to choose. For the final decision of which option to choose, the agent attributes a probability to each option based on the difference between the value of the current option and that of the lowest valued option. Options with higher differences in value have a higher probability to be chosen. This mechanism enables us to model some amount of variation in people's behavior, although this behavior is very strongly influenced by the performed planning process.

A fundamental component of this planning mechanism is the behavioral analysis that is applied to predict how another agent will regard and react to a possible interaction as well as its effects from that agent's perspective (by applying second-order Theory of Mind reasoning [54]). To perform such an analysis, the agent in the role of the actor uses its mental model of the other (target) agent and simulates the target's cognitive process to attempt to predict its social power and value assessments and, consequently, its decision.

In our model, social power strategies, when used, serve to emphasize existing social powers.[20] For instance, in the case of an expert social power, there are strategies for emphasizing the importance of the skill in the given situation, the target's trust in the actor and the difference in skill level between the target and actor. Following the same approach, we created social power strategies for all factors used in the formalization of the different bases of social power.

In summary, in the Power Interaction Planning process, the actor agent decides which is the best course of action to follow in order to influence another agent. The selected interaction mode may abstain from the use of a social power strategy or may use a specific social power strategy that emphasizes a given base of social power. To make this decision, the agent performs a situational social power analysis and considers its own agent-specific preferences for strategies relying on different bases of social power as well as their effects upon itself and the target of the influence interaction. For the pseudocode of this cognitive process, see Algorithm 10 in Appendix A.

### 5.4. Using the SAPIENT framework in different contexts

To operationalize the developed cognitive architecture, we implemented it in the SAPIENT framework. It was developed as a generic C# library targeting .NET 3.5 to maximize compatibility with Unity™, which is frequently used to build intelligent agent systems. The framework is accompanied by a set of tests that enabled us to verify the correct implementation of the social power mechanisms. Additionally, we built an authoring tool to allow easy parameterization of both the environment and the agents created with the framework. This authoring tool significantly facilitated the processes of testing the framework and implementing and refining scenarios. The developed framework, testing procedures and authoring tool are available to the community.[21]

The SAPIENT framework (and accompanying authoring tool) provides a theoretically supported and clearly structured approach for authoring social forces as drivers of agent/character behavior that is independent of context. To create SAPIENT agents for a given context, a user specifies the characteristics that parameterize their cognitive processes and will consequently generate their behaviors. The characteristics to be specified are the agents' mental representations of their environment (e.g., existing objects and their properties), their individual characteristics (e.g., capabilities, personality, and motivations) and their social reality (e.g., liking/status relationships and social groups). Once these characteristics have been authored and the corresponding configuration file has been generated,[22] it is used by the SAPIENT framework to create the agents' minds for any subsequent simulation (discrete or continuous). The link between the simulation environment and the agents' minds is established through a set of pre-defined signals that can be communicated between the framework and the simulation environment.

The agent framework is independent of context, environmental discreteness and simulation representation (e.g., 3D environment or text-based). Moreover, if there are any pre-existing simulation-logic- or visual-representation-dependent factors that have not been represented in the framework, it is possible to adapt it to handle these specific cases. This is achieved by adding new signals to enable the framework to properly link with an existing environment. It is also possible (using a plug-in architecture) to extend the native signal-handling processes (that handle social-power-related signals) with scenario-specific signal handlers for any possible signal. Similarly, it is also possible to extend the agent's decision processes with scenario-specific logic.

## 6. Case study: Social Theatre

To explore the potential of the developed cognitive architecture, we implemented the presented social power model in a game to perform a pilot user study evaluation. The study focused on assessing the model's ability to create agents that are perceived to have social power awareness and on the interaction experience that such agents provide for users.

### 6.1. Scenario description

Social Theatre is a virtual environment in which a user plays the role of the director of a theatre company (see Fig. 9). The company consists of four actor agents that were created using the SAPIENT framework to integrate social power cognitive processes into the agents' decisions. In addition to conceptualizing and maintaining a social relation with the user, each actor agent has a specific preference regarding the role that actor wishes to perform in a given play. In this environment, we created a human–agent interaction setting in which the agents are affected by the social powers associated with influence interactions with the user. The game takes the user through sequences of rehearsal/performance cycles, each representing a different play. A play is characterized by a specific set of available roles that determines the difficulty of the assignments to be made by the user. In an easy situation, the available roles perfectly match the actors' preferences. In a more difficult

[20] Power strategies are related to "Preparatory Devices" in the Power/Interaction Model [66].

[21] http://gaips.inesc-id.pt/sapient.

[22] All this is done through a specifically tailored authoring tool that has a graphical user interface to facilitate interaction and to more easily organize and display the concepts to be authored. Furthermore, the tool offers built-in help information and dependency handling/alerts. For instance, if the user eliminates a type of agent motivation and has also previously specified a goal to modify that motivation, the user is asked if he/she really wishes to eliminate the motivation because doing so will also eliminate all agents' goals that depend on it.


**Fig. 9.** Social Theatre screenshot depicting the reaction of an agent to a role assignment performed by the user.[23]

situation, only one actor preference can be satisfied by the available roles, and thus, the likelihood of several actors rejecting their assigned roles is increased. Note that each different role has an associated "time on stage". As such, beyond the actor's specific role preferences, the value of any role increases with the associated "time on stage".

In each rehearsal, the user must direct the agents by assigning them roles from among the set of available roles for the current play while simultaneously attempting to influence the agents to accept these roles. If an agent does not accept the assigned role, then that agent leaves the rehearsal and will not participate in the play, thereby diminishing its quality for the audience. To influence an agent, the user can simply rely on the social power it has acquired or can use a social power strategy to emphasize one of his/her social power bases. An interaction (with or without a strategy) is realized through a specific sentence, which the user may choose from the following:

- No strategy – Tell the actor, "Will you play this role, please?"
- Coercive strategy – Tell the actor, "If you don't accept this role now, I will only give you the [worse role name] in the next play."
- Reward strategy – Tell the actor, "If you accept this role now, I will give you the one you want in the next play."
- Legitimate strategy – Tell the actor, "I am the director; I have the right to assign you any role I see fit."
- Referent liking strategy – Tell the actor, "I really like you, and I think this is the right role for you."
- Referent status strategy – Tell the actor, "You know, I was once recognized by the crowds as their favorite actor of all time."
- Expert strategy – Tell the actor, "I have vast experience in directing plays, and you will do well in this part."

If no strategy is used, then the sentence contains no reference to any social power. All other possible sentences were created such that they emphasize a particular base of social power [66]. To create them, we used the definitions of each social power and formulated specific sentences validated to be associated with each different base of social power [60].

After the assignment and strategy choice, the agents receive their role assignments and reply (following the non-anticipated cognitive path for a target agent) based on their social power assessment of the interaction. The agents decide to either accept or refuse to participate in the play and generate appropriate replies depending on the match between their desired roles, the strength of the activated social powers and any promises or threats made by the user in previous rounds. Simultaneously, and in accordance with the results of the assignment interactions and replies, the agents derive the consequent social power effects, which result in updates to their social-power-related beliefs. Finally, after the performance is given (which occurs if at least two of the four initial actors accept their roles), an agent is chosen as the audience's favorite actor. The agents update their perceptions of the favorite actor status for the selected agent, and their acting skills increase in accordance with the importance of their performed roles as a result of their participation in the play. If three plays are canceled, the user is fired from the theatre company.

[23] Virtual environment originally developed by Serious Games Interactive: http://www.seriousgames.net/.

The user study scenario was designed to present the user with assignment choices that are initially simple and increase in difficulty as the game cycles progress. The difficulty is increased by varying the available roles such that the number of actor preferences that can be satisfied decreases with each cycle. In accordance with the "time on stage" characteristics of the different roles, each actor is affected differently with regard to the skill gained after each play and the probability of being selected as the favorite actor. Furthermore, for the purposes of the study, the game was specified to last for eight rehearsal-performance loops, which we believed would balance the need for a sufficient number of interactions to form an opinion about the characters with the desire to avoid fatiguing the participants.

### 6.2. Evaluation

The game described above was developed as an adaptation of the My Dream Theatre virtual environment [11]. It was implemented as a Unity™ application that incorporates agents created using the SAPIENT framework to add the social power mechanisms needed to guide the agents' behaviors.

In this section, we present the results of the user study conducted to collect evidence regarding the capability of users to understand the social power dynamics underlying the agents' behaviors as generated by the developed social power model as well as the impact that SAPIENT agents can have on a user's experience with a virtual environment. To this end, in addition to the described scenario, we implemented an identical scenario without the developed agent model, in which the agents' behavior was scripted using a baseline decision process. The scripted agents were created such that they could generate behavior consistent with the setting but without any social-power-based reasoning. This process was dependent on the agents' desired roles and a set of probabilities of acceptance. A 95% chance of accepting the assigned role was defined for the case in which an agent was assigned the preferred role. This acceptance rate was set below 100% to serve as a baseline for the behavior of SAPIENT agents in cases in which the user has repeatedly frustrated the agent, resulting in the possibility of refusal of even a preferred role. For undesired roles, each scripted agent was assigned an acceptance rate of 40% such that situations in which the agents would accept roles other than their preferred roles were also exhibited by the agents. These acceptance rates for the scripted agents were adjusted based on preliminary tests, which enabled us to create scripted agents (without any social-power-based reasoning) that nevertheless exhibited behaviors such that they not only accepted desired roles most of the time but also accepted undesired roles some of the time.

#### 6.2.1. Experimental design, procedure and materials

To measure the participants' awareness of the social power dynamics at work in the presented scenario, we used a repeated measures design with participant randomization and counterbalancing of the order of conditions. The experiment included two conditions: one corresponding to playing the game with agents created using the developed social power model and the other corresponding to playing the game with agents created without the developed model and instead following scripted behavior.

The experimental procedure was guided by a form that provided participants with all of the necessary information and directed them through the experiment. Additionally, the experiment had no set time limit, but participants were informed that the complete process would last for approximately 1 hour and 15 minutes. The form began by presenting the participant with several profile questions (e.g., gender, age group, occupation, etc.) followed by questions regarding personality [26] and perspective-taking ability [21]. The personality questions measured the participants' characteristics in the framework of the five-factor model [19] based on their ratings of how well each of a set of sentences applied to them on a Likert scale from 1 (very inaccurate) to 5 (very accurate). The perspective-taking ability questions were used to assess the participants' "spontaneous attempts to adopt the perspectives of other people and see things from their point of view" [21]. The results were measured based on the participants' ratings of how well each of a set of sentences applied to them on a Likert scale from $-2$ (does not describe me well) to 2 (describes me very well). Next, the participants were asked to play the game under one of the experimental conditions. After the session, the players were asked a set of questions concerning their perceptions of the agents and the overall experience. For each agent, each participant responded to the Nesler et al. [60] questionnaire, rating sentences on an 11-point Likert scale from $-5$ (strong disagreement) to 5 (strong agreement). The scale addressed in this assessment included the following sub-scales: Global Power, Resistance/Control, Compliance and Power Bases. **Global Power** measured the user's overall ability to influence the agent. **Resistance/Control** measured the user's control over the agent and the ability to overcome resistance to the user's decisions. **Compliance** measured the user's perception of the agent's compliance. **Power Bases** measured the user's social power over the agent for each of the considered social power bases. Regarding the overall experience, the participants were asked to assess the experience on the Believability and Video Game Potential Scale [38] by rating sentences on a 5-point Likert scale from $-2$ (complete disagreement) to 2 (complete agreement). This scale includes the following sub-scales: Believability, Friendliness, Interest, Intelligence and Likability. After completing the feedback for the first game session, each participant was asked to play a second game session (under the other experimental condition) and to provide feedback through the same procedure. At the end of the experiment, the participants were allowed to provide free feedback on any aspect of the experiment.

#### 6.2.2. Participants

The study included 30 participants, 24 male and 6 female. In terms of age, most participants were between 18 and 25 years old ($n = 25$), but a few were under 18 years old ($n = 2$) or between 26 and 35 years old ($n = 3$). Most of the

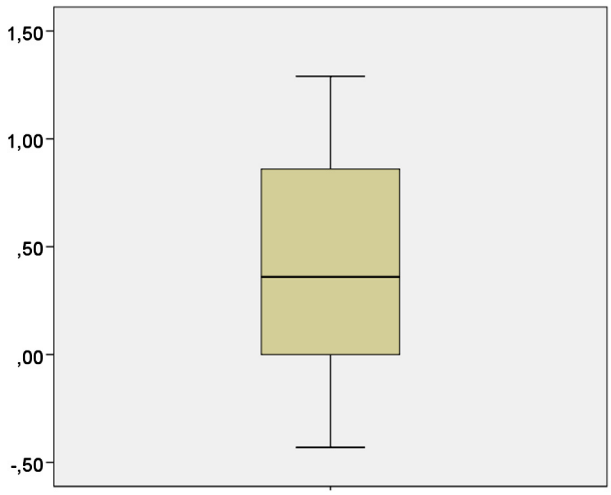


**Fig. 10.** Box plot of the participants' perspective-taking abilities.

**Table 3**
Descriptive statistics for social power awareness.

| | Scripted | | | SAPIENT model | | |
|---|---|---|---|---|---|---|
| | Median | Mean | Standard error | Median | Mean | Standard error |
| Global Power | 1.219 | 1.292 | 0.224 | **2.625** | **2.271** | 0.318 |
| Resistance/Control | 0.354 | 0.442 | 0.244 | **1.333** | **1.446** | 0.347 |
| Compliance | 1.125 | 1.171 | 0.216 | **2.500** | **2.317** | 0.340 |
| Reward Power | 1.375 | 1.197 | 0.188 | **1.625** | **1.818** | 0.276 |
| Coercive Power | 0.781 | **1.031** | 0.258 | **0.969** | 0.846 | 0.330 |
| Legitimate Power | 1.719 | 1.721 | 0.245 | **2.281** | **2.298** | 0.273 |
| Referent Power | 1.688 | 1.606 | 0.266 | **2.313** | **2.221** | 0.335 |
| Expert Power | 1.469 | 1.615 | 0.271 | **2.906** | **2.377** | 0.317 |

participants were Portuguese ($n = 28$), but there were also Brazilian ($n = 1$) and Portuguese/German ($n = 1$) participants. With regard to occupation, the participants were mostly students ($n = 27$) but also included a research assistant ($n = 1$), a software engineer ($n = 1$) and a member of the military ($n = 1$). Additionally, most participants ($n = 28$) reported using computers more than once a day, with two exceptions of people who reported using not computers but only mobile devices with that regularity.

Regarding the participants' perspective-taking abilities (see Fig. 10), they were characterized by a skew toward higher scores, with Mdn = 0.36, an upper quartile (0.36 to 0.86) slightly below 1 and a lower quartile (0 to 0.36) above or equal to 0. This means that 75% of the participants exhibited a moderate or high perspective-taking ability. The maximum ($max = 1.29$) and lower ($min = -0.43$) values reinforce the skew toward positive scores.

*6.2.3. Results*

In Table 3, we summarize the descriptive statistics for the social-power-related measurements performed for both experimental conditions. These results give a preliminary indication that the users generally perceived a higher social power awareness for the agents implemented using the SAPIENT framework.

To test this hypothesis, we performed a further statistical analysis to check whether these differences were statistically significant (see Table 4). Upon applying the Shapiro–Wilk test to the data, we found no evidence against normality for the Resistance/Control, Reward Power, Coercive Power, Legitimate Power and Referent Power dimensions. Therefore, for these dimensions, we assumed a normal distribution and applied the dependent t-test. For the remaining dimensions, the Shapiro–Wilk test showed evidence against normality and we therefore applied the Wilcoxon signed-rank test. To account for the multiple-comparison problem, we applied the Benjamini–Hochberg procedure [33] to the significance levels. For Coercive Power, Legitimate Power and Referent Power, no statistically significant differences were found. However, several significant differences between the two conditions were found, all in favor of SAPIENT agents:

- For Global Power, we found that the perception of the global power experienced by the agents that the participants reported after playing the game with the SAPIENT agents (Mdn = 2.63) was significantly higher than that reported after playing the game with the scripted agents (Mdn = 1.22). The results were $T = 3.04$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.39$.

**Table 4**
Statistical analysis of the differences between the two experimental conditions in terms of social power awareness.

| | Differences | | |
|---|---|---|---|
| | Test statistic | Significance | Effect size |
| Global Power | 3.044 | 0.002[a] | 0.393 |
| Resistance/Control | −2.479 | 0.019[a] | 0.071 |
| Compliance | 3.006 | 0.003[a] | 0.388 |
| Reward Power | −2.181 | 0.037[a] | 0.065 |
| Coercive Power | 0.530 | 0.600 | – |
| Legitimate Power | −1.733 | 0.094 | – |
| Referent Power | −1.650 | 0.110 | – |
| Expert Power | 2.870 | 0.004[a] | 0.371 |

[a] Benjamini–Hochberg (BH) significant (FDR = 0.1).

**Table 5**
Descriptive statistics for participant experience.

| | Scripted | | | SAPIENT model | | |
|---|---|---|---|---|---|---|
| | Median | Mean | Standard error | Median | Mean | Standard error |
| Believability | 0.429 | 0.349 | 0.126 | **1.000** | **0.881** | 0.128 |
| Friendliness | 0.000 | −0.060 | 0.153 | **1.000** | **0.687** | 0.157 |
| Interest | 0.000 | −0.140 | 0.213 | **0.750** | **0.643** | 0.151 |
| Intelligence | 0.000 | 0.433 | 0.213 | **1.000** | **0.600** | 0.233 |
| Likability | 0.000 | −0.273 | 0.223 | **1.000** | **0.660** | 0.157 |

**Table 6**
Statistical analysis of the differences between the two conditions in terms of participant experience.

| | Differences | | |
|---|---|---|---|
| | Test statistic | Significance | Effect size |
| Believability | −3.879 | 0.001[a] | 0.088 |
| Friendliness | 3.130 | 0.002[a] | 0.404 |
| Interest | 3.038 | 0.002[a] | 0.392 |
| Intelligence | 0.554 | 0.580 | – |
| Likability | 2.939 | 0.003[a] | 0.379 |

[a] Benjamini–Hochberg (BH) significant (FDR = 0.1).

- For Resistance/Control, we found that on average, the perception of the underlying social power force mechanism that the participants reported after playing the game with the SAPIENT agents ($M = 1.45$, SE $= 0.35$) was significantly higher than that reported after playing the game with the scripted agents ($M = 0.44$, SE $= 0.24$). The results were $t(29) = -2.48$ for the dependent t-test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = .07$.
- For Compliance, we found that the perception of agent compliance that the participants reported after playing the game with the SAPIENT agents (Mdn $= 2.50$) was significantly higher than that reported after playing the game with the scripted agents (Mdn $= 1.13$). The results were $T = 3.01$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.39$.
- For Reward Power, we found that on average, the perception of the reward power experienced by the agents that the participants reported after playing the game with the SAPIENT agents ($M = 1.82$, SE $= 0.28$) was significantly higher than that reported after playing the game with the scripted agents ($M = 1.20$, SE $= 0.19$). The results were $t(29) = -2.18$ for the dependent t-test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.07$.
- For Expert Power, we found that the perception of the expert power experienced by the agents that the participants reported after playing the game with the SAPIENT agents (Mdn $= 2.91$) was significantly higher than that reported after playing the game with the scripted agents (Mdn $= 1.47$). The results were $T = 2.87$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.37$.

In this study, we also measured the participants' experience with the agents. In Table 5, we summarize the descriptive statistics for the experience-related measurements performed for both experimental conditions. These findings give a preliminary indication that the users rated the agents implemented with the SAPIENT framework more highly in all dimensions.

To test this hypothesis, we performed a further statistical analysis to check whether these differences were statistically significant (see Table 6). Upon applying the Shapiro–Wilk test to the data, we found no evidence against normality for Believability. For this dimension, we therefore assumed a normal distribution and applied the dependent t-test. For the remaining dimensions, the Shapiro–Wilk test showed evidence against normality and we therefore applied the Wilcoxon signed-rank test. To account for the multiple-comparison problem, we applied the Benjamini–Hochberg procedure [33]

**Table 7**
Table of correlations between the social-power-base-specific measurements (from the social power awareness analysis) and the employed power strategies under the SAPIENT condition. Abbreviations: NoS – no strategy; RewardS – reward strategy; CoerciveS – coercive strategy; LegitimateS – legitimate strategy; ReferentS – referent strategy; ExpertS – expert strategy.

| | NoS | RewardS | CoerciveS | LegitimateS | ReferentS | ExpertS |
|---|---|---|---|---|---|---|
| Reward Power | −0.49[a] | 0.56[a] | −0.11 | −0.06 | 0.43[a] | 0.19 |
| Coercive Power | −0.19 | −0.07 | 0.16 | 0.32 | 0.09 | 0.28 |
| Legitimate Power | −0.42[a] | 0.37[a] | −0.21 | −0.14 | 0.40[a] | 0.29 |
| Referent Power | −0.50[a] | 0.55[a] | −0.11 | −0.02 | 0.43[a] | 0.23 |
| Expert Power | −0.26 | 0.26 | −0.14 | −0.14 | 0.28[a] | 0.19 |

[a] Benjamini–Hochberg (BH) significant (FDR = 0.1).

to the significance levels. For Intelligence, no statistically significant differences were found. However, we found several significant differences between the two conditions for the other dimensions:

- For Believability, we found that on average, the perception of the agents' believability that the participants reported after playing the game with the SAPIENT agents ($M = 0.881$, SE $= 0.128$) was significantly higher than that reported after playing the game with the scripted agents ($M = 0.349$, SE $= 0.126$). The results were $t(29) = -3.879$ for the dependent t-test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.088$.
- For Friendliness, we found that the perceived friendliness of the agents that the participants reported after playing the game with the SAPIENT agents (Mdn $= 1.000$) was significantly higher than that reported after playing the game with the scripted agents (Mdn $= 0{,}000$). The results were $T = 3.130$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.404$.
- For Interest, we found that the perceived interest of the agents that the participants reported after playing the game with the SAPIENT agents (Mdn $= 0.750$) was significantly higher than that reported after playing the game with the scripted agents (Mdn $= 0.000$). The results were $T = 3.038$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.392$.
- For Likability, we found that the perceived likability of the agents that the participants reported after playing the game with the SAPIENT agents (Mdn $= 1.000$) was significantly higher than that reported after playing the game with the scripted agents (Mdn $= 0.000$). The results were $T = 2.939$ for the Wilcoxon signed-rank test, Benjamini–Hochberg significant ($FDR = 0.1$) with an effect size of $r = 0.379$.

Additionally, to better understand the participant's characterization of social power as either a more abstract concept or specifically related to different bases of social power, we performed a statistical analysis of the data related to the agents implemented with SAPIENT to explore the relation between the assessment of the different bases of social power and the social power strategies employed. The matrix of correlations between all of the social-power-base dimensions and the different social power strategies is shown in Table 7. Based on the previous normality analysis, there was no evidence against normality for Reward Power, Coercive Power, Legitimate Power or Referent Power. Additionally, upon applying the Shapiro–Wilk test to the data regarding the different social power strategies employed, we found no evidence against normality for any of them. For correlations including the Expert Power dimension, we used the Kendall correlation method [33]. For the remaining correlations, we used the Pearson correlation method. To account for the multiple-comparison problem, we applied the Benjamini–Hochberg procedure [33] to the significance levels.

*6.2.4. Discussion*

In terms of social power awareness (see Table 3), we may generally observe that with the exception of the perceived Coercive Social Power, all median and mean values are always higher for the condition corresponding to the agents implemented using the developed social power model. If we compare the mean Coercive Power values between the two conditions, we observe a difference opposite to that expected. However, this difference is very small, and further statistical analysis showed that it is not (or even close to) statistically significant. One reason for this is that the users avoided using coercion; on average, fewer than 2% of their interactions used a coercive social power strategy. Regarding the measurements for Legitimate and Referent Power, although the differences between the means support our hypothesis, after subsequent analysis, we found no statistically significant differences (see Table 4). To resolve these inconclusive results in a future study, we can improve our design in two ways. First, we can increase the number of rehearsal-performance loops to provide participants with more interaction time with the agents and enable the exploration of more interactions. This is especially relevant for coercive social power because its use requires executing and monitoring the results of the associated coercive action and thus spans more than one round. Second, we can also increase the number of participants, thereby increasing the sample size and thus the likelihood of finding significant differences between the two conditions.

Nonetheless, we also measured several significant differences between the two conditions that support our hypothesis of increased perceived social power awareness in the agents implemented using the developed social power model. The differences measured for Global Power, Compliance and Expert Power (see Table 4) were not only significant but also associated

with a moderate effect size.[24] We believe that significant effects were observed in these measurements because of their broad range (Global Power and Compliance) and salience to the scenario (Expert Power) compared with the other dimensions. The Expert Power dimension is strongly associated with skill, which was one of the most salient characteristics to the user in the presented scenario. The differences measured for the Resistance/Control and Reward Power dimensions were also significant but associated with a small effect size. We believe that these differences can be attributed to different reasons. Measurement of the Resistance/Control dimension depends on the perception of conflicting desires and social-power forces in the agents, which was not very salient in the assignment or reaction interactions. Meanwhile, the small effect observed for Reward Power may be explained by the fact that the interaction dynamics span more than one interaction, thus making the effect of this power more difficult to perceive.

Based on the measurements related to the different bases of social power, we also explored the participants' perceptions of the different bases, namely, whether there was evidence that social power was perceived as a single construct or as being composed of multiple different concepts. To look for evidence regarding the perception of the different bases of social power, we explored the correlations between the ratings of the different bases of social power and the social power strategies employed by the users (under the SAPIENT condition). Analysis of Table 7 yielded mixed results, some pointing to a single-construct perception and others suggesting some level of differentiation. The use of no strategy is negatively correlated with all social-power-base dimensions, and significantly so, with the exception of the Coercive and Expert dimensions. This pattern is consistent with our expectations because when no strategy is applied, no social powers are made salient. The reward and referent strategies show significant positive correlations (and, in each case, the highest among all correlations observed for that strategy) with their associated bases of social power, as expected from the fact that each strategy involves emphasizing the base of social power with which it is associated. This offers some indication that the participants perceived these bases of social power differently through the different social power strategies available. Similar evidence (though less strong) is seen for the expert strategy, which also has a positive correlation with the Expert Power dimension; however, this correlation is neither significant nor the highest among all correlations observed for this strategy. By contrast, for the coercive and legitimate strategies, there are no significant correlations matching these strategies with their associated bases of social power as expected; therefore, these dimensions offer no support for the users' differentiation of the different bases of social power. For the coercive strategy, there is also no significant correlation, although it is encouraging to note that among all correlations for this strategy, only that with the Coercive Power dimension is positive. It is possible that in a future study, by increasing the number of interactions, we could extract a significant correlation for this dimension. The results of analyzing the legitimate strategy are inconsistent with our expectations; we expect this dimension to be positively correlated with Legitimate Power, but in fact, it is non-significantly and negatively correlated with Legitimate Power. Furthermore, the legitimate strategy has an unexpected significant positive correlation with Coercive Power. A possible explanation for this is that the coercive base of social power is, in fact, intrinsically related to the legitimate base of social power (as described in Section 4.3.2). This might indicate that the legitimate strategy created for the scenario exhibits a strong relation to the co-existing coercive social power. One important issue regarding the performed analysis of the coercive and legitimate strategies is that these strategies were by far the least used by the participants (2% and 3%, respectively, compared with 23% for the referent strategy, 13% for the reward strategy and 11% for the expert strategy).

Although some strategy-related dimensions are significantly and positively correlated with the expected social-power-base dimensions and suggest that the users' perceived the existence of the different bases of social power, several other results do not show the expected correlations and do not offer evidence of the expected separation of concepts. Overall, we obtained some results that indicate some degree of separation the social power bases, but there is also significant evidence to the contrary. To directly address this issue in a future study, we can individually manipulate each base of social power to more clearly ascertain the participants' ability to distinguish between the different bases of social power modeled in SAPIENT.

Overall, the results for social power awareness support the value of the developed social power model and its ability to simulate humanlike social power dynamics. The majority of the results are consistent with our hypothesis, and for most of the dimensions measured, we also found significant differences and effect sizes to support it. The only exception was the Coercive Power dimension, but the differences measured between the conditions were very small and far from statistically significant.

Regarding the participants' experience with the agents (see Table 5), we observe that for all measured dimensions, the median and mean values are always higher for the SAPIENT agents. A more in-depth statistical analysis of these differences yielded a better understanding of these results. Regarding Intelligence, although the participants rated the SAPIENT agents higher on average, this difference was not significant. We believe that this result can be attributed to the fundamental attribution error[25] (FAE) on the part of several participants. Because the participants were given no information regarding the agents' decision models, many may have attributed some of the less coherent or expected behaviors to personality characteristics (although the agents were all equally parameterized) instead of situational factors.

[24] A small effect corresponds to $r = 0.10$; a moderate effect, to $r = 0.30$; and a large effect, to $r = 0.50$ [32].

[25] The fundamental attribution error is "the tendency for people to over-emphasize personality-based explanations for behaviors observed in others while under-emphasizing the role and social power of situational influences on the same behavior" [76].

Nonetheless, for the remaining dimensions, we measured significant differences between the two conditions, supporting our hypothesis of a better interaction experience under the SAPIENT condition. The differences measured for Friendliness, Interest and Likability (see Table 6) were significant and associated with a moderate effect size. We believe that these effect sizes can be attributed to the players' expectations regarding their social interactions with the agents in combination with the more socially coherent behaviors of the agents implemented using the developed social power model. This reasoning is supported by comments made by the participants during the experiment that indicated some frustration with unexpected behaviors demonstrated by the agents implemented using the scripted decision model. The differences measured for Believability were also significant but associated with a small effect size. We believe that the reason for this finding also lies in the differences between the agent implementations, combined with the FAE on the part of the participants. If participants attributed less coherent behaviors to personality traits, then they might have still found the agents to be, to some extent, believable.

Overall, we found strong evidence that the SAPIENT agents were more positively regarded than those following the scripted model. All results obtained regarding the interactive experience are consistent with our hypothesis, and for all dimensions except Intelligence, we also found significant differences and effect sizes to support it.

Although social power is very pervasive, it is usually an implicit concept manipulated through many other social concepts, unless it is directly referenced. Additionally, considering the careful design of the experiment (virtual environment setting, materials, procedure) and the careful implementation of the agents with the scripted decision model, we believe that we were successful in minimizing possible biases that might have resulted from the participants identifying the subject under assessment in the experiment. This argument is supported by the observation of several positive ratings throughout the measurements for the scripted agents, although they were generally scored lower than the agents implemented using the developed social power model. Interestingly, in qualitative feedback given by several participants after the experiment, they specifically stated they could not identify what was being studied but, in several instances, mentioned that one set of agents (the SAPIENT agents) appeared more coherent, although they could not explain why. As such, although the participants did not know what was being assessed nor, in many instances, could pinpoint a specific difference between the conditions, it is clear that the agents implemented using the SAPIENT model created the perception of strong social power dynamics, resulting in stronger and more positive effects in terms of the interactive experience.

Even so, it might be argued that because the scripted agents received positive ratings, using such agents when implementing agents for social settings is a valid approach requiring much less effort. We do not agree for several reasons. First, we believe that such positive ratings are closely bound to the required degree of dynamic interaction generation. In the presented scenario, the possible interactions were achieved through selection among several sentences specifically composed for the situation. If we had presented a scenario with many more possible situations that required dynamic interaction generation coherent with some model of behavior, then the generation of suitable interaction configurations would quickly become impractical because all potential interactions would have to be individually configured. By contrast, for agents implemented using the SAPIENT model, the gradient of behaviors exhibited in different situations is already dynamically generated by the underlying social power mechanisms in accordance with the manipulated social concepts. Second, and possibly an even stronger argument, is that users interacting with agents can always make unexpected attributions to them. If we use the SAPIENT model, we can parameterize a given behavior to achieve a more valid simulation of it, whereas in the case of scripted agents, the users' attributions may frequently be unrelated to the intended model of behavior.

#### *6.2.5. Participants' performance and utilization of social power strategies*

During the user study using the Social Theatre environment, we collected various data that enabled us to analyze the participants' choices regarding social power strategies and their impact on the participants' performance in the game. Because different sets of data were required for this analysis, in Fig. 11, we show the distribution of the players' choices in all interactions (on the left), the numbers of desired role assignments made by the participants (on the right), and the outcomes of the interactions between the study participants and the actor agents (measured in terms of the number of agents participating in the plays).

Based on Fig. 11, we wish to highlight some interesting cases.[26] Participant 1 never assigned the agent actors the roles they desired but nevertheless achieved participation rates near 90%. However, inspecting the strategies used by this participant suggests that these results probably derived from the intensity and variety of strategy use compared with the majority of the other users.

Similarly, although they did not assign an equally high number of undesired roles, participants 3 and 17 also assigned a high number of undesired roles to actor agents. Similarly, they also used a large number and variety of strategies, although not as high as participant 1. Their more agreeable behavior toward the actors' desired roles combined with a more moderate use of social power strategies produced better/similar results in terms of participation. By contrast, participant 5 assigned the highest percentage of desired roles, achieving a participation rate of 100% while relying less on power strategies.

The lowest participation rates of slightly above 70% were achieved by participants 24, 27 and 28. These results may be explained by poor use of power strategies and by low rates of desired role attribution, especially for participants 24 and 27. The results highlight the importance of using strategies, emphasized by the aforementioned asymmetries. In fact,

[26] To relate the strategies used with the participants' individual ratings, please see Table B.9 in Appendix B.

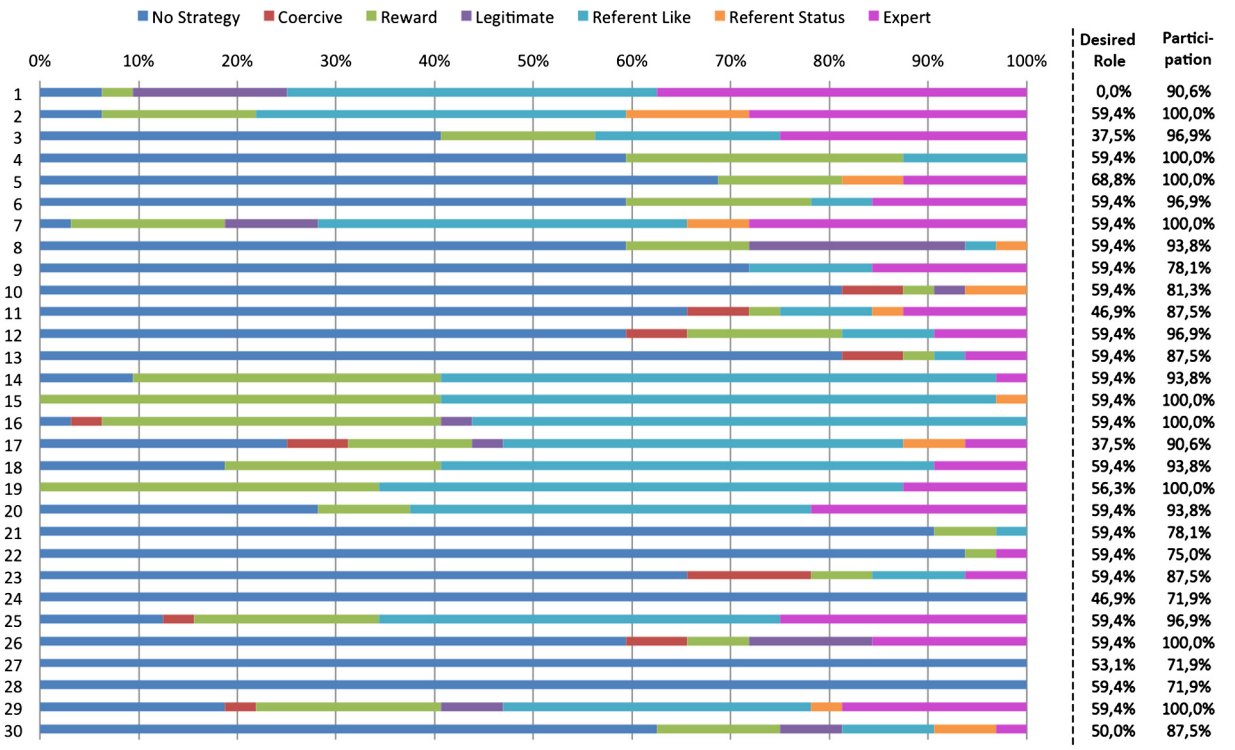


**Fig. 11.** Distribution graph of the participants' interaction strategies (on the left) and percentages of desired role assignments made and agents participating in the plays (on the right).

participants 15 and 19, who achieved 100% agent participation, always relied on power strategies, especially reward and referent power strategies. Furthermore, participant 14 also made heavy use of power strategies, achieving a near-100% participation rate.

Overall, the utilization of social power strategies seemed to have the participants' desired effect of influencing the agents to participate in the plays, even in situations in which they were assigned undesired roles.

To test this hypothesis, we analyzed the statistical correlations between the percentages of the interactions in which the participants used social power strategies and their overall performances as measured in terms of the participation percentage. The results show that the two dimensions are strongly positively correlated: $r(28) = 0.768$, $p < 0.001$.

These results empirically indicate that the utilization of social power strategies allowed the users to achieve good results even in situations in which the goals (desired and undesired roles) of the actor agents offered resistance to the acceptance of the roles assigned by the players. This demonstrates that the model generates believable behavior and enables the creation of interesting settings (from a game design perspective) for applications such as serious games.

### 6.2.6. *Performance of the SAPIENT framework*

An important issue when using frameworks that implement complex cognitive processes and recursive representations (e.g., Models of Others/Theory of Mind) is the performance of the system. If the social power processes of SAPIENT were computationally costly, this could hinder its utilization in complex settings with many agents. To address this issue and determine the potential impact of the core social power processes in an agent implementation using SAPIENT, we evaluated the performance of these processes. To this end, we ran a set of simulations using a scenario identical to Social Theatre. The only difference was that the role of the company director was played by a SAPIENT agent instead of a person, as in the user study.

We created a base director parameterization based on personality data collected from each participant in the user study. From these base parameterizations, we created complete parameterizations by varying the agent's achievement, affiliation and status motivation values, characteristics for which we did not collect data in the user study. Each of the explored characteristics was represented by a probability-like variable. To achieve complete coverage of the parameter range, we chose a granularity of 0.25 for exploring different motivation values. Because we had 30 user profiles, we simulated $NSimulations(0.25, 30) = 3750$ parameterizations (see Formula (15)).

$$NSimulations(granularity, profiles) = \Bigg( \underbrace{1 + \frac{1}{granularity}}_{\substack{\#\text{ of combinations} \\ \text{per motivation}}} \Bigg)^{3} * profiles \tag{15}$$

**Table 8**
Performance assessment of core social power processes.

| | Number of calls | Minimum | Maximum | Average | Standard deviation |
|---|---|---|---|---|---|
| Power Situational Analysis | 13107688 | <1 ms | 115 ms | 0.023 ms | 0.629 ms |
| **Power Effect Assessment** | 19449033 | <1 ms | 4914 ms | 0.085 ms | 1.863 ms |
| **Power Interaction Planning** | 120558 | <1 ms | 121 ms | 0.779 ms | 3.429 ms |

The performance assessment of the core social power processes is presented in Table 8. Overall, we observe that none of the three core social power processes (including all required Theory of Mind reasoning) is computationally costly (even within two standard deviations). However, among the maximum execution times, several very high values are observed compared with the averages. These values may have occurred because of the lack of optimization in the code to specifically cope with, for example, the possible occurrence of garbage collection during the execution of these processes. Certainly, if we create scenarios in which the number of agents and the complexity of the scenario are dramatically increased, the performance is bound to diminish. However, the average execution time for each of these social power processes is very low,[27] which means that the system can clearly be scaled to scenarios involving dozens more agents without hindering its utilization, even in real-time applications.

## 7. Conclusions

Regardless of significant advances in intelligent agents' social capabilities over the past decade, there remains a significant gap in their social intelligence that seriously limits the range of social phenomena that can be simulated. To address this gap in social intelligence and, consequently, in the intelligence, expressiveness and believability of agents, we studied the links between different human cognitive processes, the different bases of social power and their underlying factors. Based on fundamental insights from the social sciences, in which social power has long been studied, we introduced a mechanism for an agent's decision-making processes incorporating five bases of social power. This mechanism was then integrated into a cognitive architecture for social-power-intelligent agents following an integrative conceptualization of social power. The modeled bases of social power are those of reward, coercive, legitimate, referent and expert power. Social power awareness with regard to different bases of social power enables agents to manipulate several social concepts related to important social processes and consequently enables more socially situated intelligent interactions between agents and users.

To describe the proposed social power mechanism, we discussed each base of social power in detail, from their definition to their dynamics in a social environment. By doing so, we clearly identified the types of situations and the range of behaviors that can be represented. Additionally, when conceptualizing the developed agent architecture, we identified three core social power processes ("Power Situational Analysis", "Power Effect Assessment" and "Power Interaction Planning") and described their integration with the typical agent processes of sensing, reasoning and interaction in a given environment. Beyond the core processes required to operationalize social power in an agent environment, the developed social power model enabled us to integrate several social constructs that can be used to parameterize agent social behavior. Of particular interest is the relation between personality and individualized agent behaviors, motivations for diverse motivational forces and separate mental models associated with Theory of Mind capabilities. A key contribution of this work is the integration of the conceptualization of social power with social psychology models, enabling us to measure the characteristics of individuals through validated questionnaires such that these measurements can be subsequently used to parameterize our model. As a result, the developed framework provides a clearly structured approach for authoring social simulations based on characteristics that not only are readily identifiable but can be measured in people.

Based on the developed architecture, we implemented the SAPIENT agent framework to create social-power-intelligent behavior in a context-independent manner, thereby facilitating the adaptation of the agents to different situations. Using this framework, we created a virtual environment to explore a setting for human–agent interaction that enabled us to perform a user study to assess users' perceptions regarding the social power awareness of SAPIENT agents and their experience in interacting with such agents. As a result, we found evidence supporting the hypothesis that the developed social power model is valuable for the development of agent systems that require social-power-intelligent behaviors. Additionally, we found evidence indicating that agents implemented using the developed model provide a positive interaction experience; in particular, they exhibit higher believability compared with scripted agents (unrelated to any insights from the social sciences). However, some measurements were inconclusive, and we therefore proposed extensions to the experimental design to strengthen the current results. Furthermore, as a future direction of research to consolidate the evidence from the pilot user study, we believe that a new study consisting of a Wizard of Oz experiment should be conduced to compare the behavior of SAPIENT agents with human behavior.

One future extension of the presented work would be to expand the effect mechanisms to more extensively cover factor-specific dynamics (e.g., norm dynamics) that were beyond the scope of the current study but have an important impact on social power dynamics. Another future possibility would be to extend the model to consider the effects of multiple influencing actors in a given situation. Nevertheless, the mechanisms and framework described in the present paper

[27] For example, compatible with the resources that are typically available for AI in games [59].

can already be used to improve agents' intelligence in a variety of contexts and applications. Our goal is to use the developed agent model in several settings. A first application will be the integration of the framework into a component-based asset repository for inclusion in applied games for the training of social skills.[28] The model will thus be used in interactive virtual agent-based settings to improve training applications (e.g., serious games for leadership) and enhance social simulation games[29] (e.g., role-playing or simulation games with more dynamic character social interactions). Other future applications include multi-agent simulations to study social power dynamics as a function of variations in agents' personal motivations, their personalities and the structure of the social environment in which they interact.

## Acknowledgements

This work was supported by national funds provided through Fundação para a Ciência e a Tecnologia (FCT) (UID/CEC/50021/2013) and an FCT scholarship (SFRH/BD/66663/2009). This work was also supported by the INVITE project (UTA-Est/MAI/0008/2009) funded by FCT under a UT-Austin/Portugal cooperation agreement.

## Appendix A. Pseudocodes for the core social power processes

The pseudocodes for the core social power processes of the SAPIENT framework: Power Situational Analysis (see Algorithm 1), Power Effect Assessment (see Algorithm 7) and Power Interaction Planning (see Algorithm 10).

**Algorithm 1** Power Situational Analysis pseudocode for the identification of the social powers that are active in a (potential) interaction situation.

```
1: procedure PowerSituationalAnalysis(actor, target, behavior, strategy)
2:     isp ← []                                   ▷ Structure to contain all identified social powers.
3:     ach ← AchievementMotivation(target)
4:     dut ← DutifulnessMotivation(target)
5:     aff ← AffiliationMotivation(target)
6:     sta ← StatusMotivation(target)
7:     ▷ Identification of reward social powers
8:     isp ← PowerSituationalAnalysisReward(actor, target, behavior, strategy, isp, ach)
9:     ▷ Identification of coercive social powers
10:    isp ← PowerSituationalAnalysisCoercive(actor, target, behavior, strategy, isp, ach)
11:    ▷ Identification of legitimate social powers
12:    isp ← PowerSituationalAnalysisLegitimate(actor, target, behavior, strategy, isp, dut)
13:    ▷ Identification of referent social powers
14:    isp ← PowerSituationalAnalysisReferent(actor, target, behavior, strategy, isp, aff, sta)
15:    ▷ Identification of expert social powers
16:    isp ← PowerSituationalAnalysisExpert(actor, target, behavior, strategy, isp)
17:    isp ← ApplyStrategy(isp, strategy)         ▷ Enhances the emphasized social power factor.
18:    return isp
```

**Algorithm 2** Power Situational Analysis pseudocode for the identification of reward social powers.

```
1: procedure PowerSituationalAnalysisReward(actor, target, behavior, strategy, isp, ach)
2:     for all rt ∈ RewardTendencies(actor) do      ▷ Identification of reward social powers.
3:         if AssociatedBehavior(rt) = behavior then
4:             ra ← RewardBehavior(rt)
5:             rav ← Value(ra)
6:             rtv ← TendencyValue(rt)
7:             isp ← RewardSocialPower(rav, ach, rtv)
8:     return isp
```

**Algorithm 3** Power Situational Analysis pseudocode for the identification of coercive social powers.

```
1: procedure PowerSituationalAnalysisCoercive(actor, target, behavior, strategy, isp, ach)
2:     for all ct ∈ CoerciveTendencies(actor) do
3:         if AssociatedBehavior(ct) = behavior then
4:             ca ← CoerciveBehavior(ct)
5:             cav ← Value(ca)
6:             ctv ← TendencyValue(ct)
7:             isp ← CoerciveSocialPower(cav, ach, ctv)
8:     return isp
```

[28] Realising an Applied Gaming Eco-system – http://rageproject.eu/.

[29] A subgenre of simulation-based entertainment games that simulate individuals and social interactions to some extent (e.g., The Sims game series).

**Algorithm 4** Power Situational Analysis pseudocode for the identification of legitimate social powers.

```
1: procedure PowerSituationalAnalysisLegitimate(actor, target, behavior, strategy, isp, dut)
2:     for all sg ∈ SocialGroups(target) do
3:         if actor ∈ GroupMembers(sg) then
4:             for all n ∈ Norms(sg) do
5:                 gi ← GroupImportance(target, sg)
6:                 gc ← GroupConformity(target, sg)
7:                 ns ← GetNormSalience(n)
8:                 switch NormType(n) do
9:                     case position
10:                        if PlaysActorRole(n, actor) & PlaysTargetRole(n, target) then
11:                            isp ← LegitimateSocialPower(gi, dut, gc, ns)
12:                    case actioncommitment
13:                        if ActionCommitment(target, actor, behavior) ∈ Commitments(target) then
14:                            isp ← LegitimateSocialPower(gi, dut, gc, ns)
15:                    case reciprocity
16:                        hea ← HelpfulActionsRatio(target, sg)
17:                        isp ← LegitimateSocialPower(gi, dut, gc, ns, hea)
18:                    case equity
19:                        haa ← HarmfulActionsRatio(target, sg)
20:                        isp ← LegitimateSocialPower(gi, dut, gc, ns, haa)
21:                    case responsibility
22:                        de ← Dependence(actor, target)
23:                        isp ← LegitimateSocialPower(gi, dut, gc, ns, de)
24:    return isp
```

**Algorithm 5** Power Situational Analysis pseudocode for the identification of referent social powers.

```
1: procedure PowerSituationalAnalysisReferent(actor, target, behavior, strategy, isp, aff, sta)
2:     if LikingRelation(target, actor ∈ LikingRelations(target) then
3:         lv ← LikingValue(target, actor)
4:         isp ← ReferentSocialPower(lv, aff)
5:     for all sc ∈ AssociatedStatusCategories(behavior) do
6:         if sc ∈ StatusCategoriesPreferences(target) then
7:             sp ← StatusPreferenceValue(target, sc)
8:             sv ← StatusValue(target, actor, sc)
9:             isp ← ReferentSocialPower(sp, sta, sv)
10:    return isp
```

**Algorithm 6** Power Situational Analysis pseudocode for the identification of expert social powers.

```
1: procedure PowerSituationalAnalysisExpert(actor, target, behavior, strategy, isp)
2:     for all sk ∈ AssociatedSkills(behavior) do
3:         sa ← SkillLevel(actor)
4:         st ← SkillLevel(target)
5:         if sa > st then
6:             si ← SkillImportance(sk)
7:             tr ← TrustValue(target, actor)
8:             isp ← ExpertSocialPower(si, tr, sa, st)
9:     return isp
```

**Algorithm 7** Power Effect Assessment pseudocode.

```
1: procedure PowerEffectsAssessment(actor, target, isp, strategy, accepted)
2:     if accepted = true then
3:         PowerEffectsAssessmentAccept(actor, target, isp, strategy)
4:     else
5:         PowerEffectsAssessmentRefuse(actor, target, isp)
```

**Algorithm 8** Power Effect Assessment pseudocode for the case in which the influence interaction was successful.

```
1: procedure PowerEffectsAssessmentAccept(actor, target, isp, strategy)
2:     st ← StrategyType(strategy)
3:     ef ← EmphasizedFactor(strategy)
4:     for all sp ∈ isp do
5:         switch SocialPowerType(sp) do
6:             case reward                                   ▷ Reward social power (accept) effects.
7:                 if st = reward & ef = RewardsSusceptibility then
8:                     Increase(actor, RewardsSusceptibility, target)
9:                 if ReceivedReward(target = true) then
10:                    Increase(target, Liking, actor)
11:                    if st = reward & ef = RewardLikelihood then
12:                        Increase(target, RewardLikelihood, actor)
13:                else
14:                    Decrease(target, RewardTendency, actor)
15:            case coercive                                 ▷ Coercive social power (accept) effects.
16:                if st = coercive & ef = CoercionsSusceptibility then
17:                    Increase(actor, CoercionsSusceptibility, target)
18:                if st = coercive & ef = CoercionTendency then
19:                    Increase(target, CoercionTendency, actor)
20:                ca ← CoerciveBehavior(sp)
21:                if HasNormativeSupport(ca) = false then
22:                    Decrease(target, Liking, actor)
23:            case legitimate                               ▷ Legitimate social power (accept) effects.
24:                if st = legitimate & ef = ActorNormConformity then
25:                    Increase(target, NormConformity, actor)
26:                if st = legitimate & ef = TargetNormConformity then
27:                    Increase(target, NormConformity, target)
28:            case referent                                 ▷ Referent social power (accept) effects.
29:                sut ← SubType(sp)
30:                if sut = Liking & st = referent & ef = LikingRelation then
31:                    Increase(target, Liking, actor)
32:                if sut = Status & st = referent & ef = StatusRelation then
33:                    Increase(target, StatusRecognition, target)
34:            case expert                                   ▷ Expert social power (accept) effects.
35:                if st = expert & ef = TrustRelation then
36:                    Increase(target, Trust, actor)
37:                if st = expert & ef = LowTargetSkill then
38:                    Decrease(target, SkillLevel, target)
39:                if st = expert & ef = HighActorSkill then
40:                    Increase(target, SkillLevel, actor)
```

**Algorithm 9** Power Effect Assessment pseudocode for the case in which the influence interaction was unsuccessful.

```
1: procedure PowerEffectsAssessmentRefuse(actor, target, isp)
2:     for all sp ∈ isp do
3:         switch SocialPowerType(sp) do
4:             case reward                                   ▷ Reward social power (refuse) effects.
5:                 Decrease(actor, RewardsSusceptibility, target)
6:             case coercive                                 ▷ Coercive social power (refuse) effects.
7:                 Decrease(actor, CoercionsSusceptibility, target)
8:                 ca ← CoerciveBehavior(sp)
9:                 if HasNormativeSupport(ca) = false then
10:                    Decrease(target, Liking, actor)
11:                if ReceivedCoercion(target = true) then
12:                    Increase(target, CoercionTendency, actor)
13:                else
14:                    Decrease(target, CoercionTendency, actor)
15:            case legitimate                               ▷ Legitimate social power (refuse) effects.
16:                Decrease(actor, NormConformity, target)
17:            case referent                                 ▷ Referent social power (refuse) effects.
18:                sut ← SubType(sp)
19:                if sut = Liking then
20:                    Decrease(target, Liking, actor)
21:                if sut = Status then
22:                    Decrease(target, StatusRecognition, target)
23:            case expert                                   ▷ Expert social power (refuse) effects.
24:                Decrease(target, Trust, actor)
25:                Decrease(target, SkillLevel, actor)
```

**Algorithm 10** Power Interaction Planning pseudocode for the case in which an influence interaction is required.

```
1: procedure POWER INTERACTION PLANNER(actor, target, behavior)
2:     op ← []
3:     ▷ Interaction strategies range from no strategy to strategies specific to each base of social power.
4:     for all strategy ∈ InteractionOptions do
5:         isp ← PowerSituationalAnalysis(actor, target, behavior, strategy)
6:         va ← ValueAssessment(target, behavior)
7:         ▷ The effect prediction procedure simulates the decision of the target based on social power and value assessments. The effects are predicted in a manner similar to Algorithm 7 but are not applied.
8:         ef ← PredictEffects(sip, va, strategy)
9:         ▷ The predicted effects are assessed in terms of value with regard to the motivations and goals of the agent.
10:        ev ← ValueAssessment(ef)
11:        bi ← PreferenceBias(actor)
12:        ▷ A preference bias increases the evaluation of one or more associated strategies.
13:        bi ← BiasedInfluenceForce(isp, bi)
14:        op ← va + bi + ev
15:    ▷ The selection procedure stochastically selects an option based on a probability value for each option equal to the percentage of the summed values for all options represented by the value for the associated option.
16:    return Select(op)
```

## Appendix B. Social power measurements: ratings of individual participants

**Table B.9**
Individual ratings for all participants in the pilot user study with regard to the social power measurements performed.

| Participant | Global power | Resistance/Control | Compliance | Reward | Coercive | Legitimate | Referent | Expert |
|---|---|---|---|---|---|---|---|---|
| 1 | 3.25 | 1.75 | 1.88 | 0.9 | 2.31 | 1.63 | 1.25 | 2.13 |
| 2 | 4.19 | 4.21 | 4.5 | 4.25 | 1.56 | 4.56 | 3.94 | 3.88 |
| 3 | 3.38 | 3.63 | 3.63 | 3.3 | −2.81 | 2.75 | 3.94 | 4 |
| 4 | 3.63 | 2.58 | 4.63 | 3 | −2.13 | 3.19 | 4.94 | 2.81 |
| 5 | 2.13 | 1.88 | 3.75 | 1.55 | 2.25 | 1.88 | 1.38 | 1.75 |
| 6 | 3.13 | 3.63 | 4.13 | 3.4 | 4.06 | 3.13 | 3.38 | 3.38 |
| 7 | 4.75 | 2.92 | 4.38 | 3.85 | 2.94 | 4 | 3.88 | 3.38 |
| 8 | 1.81 | 1.25 | 1.13 | 0.65 | 1 | 0.63 | 1.13 | 0.69 |
| 9 | 0.75 | 0.42 | 3.38 | 1.4 | 1.31 | 4.19 | 3.75 | 4.31 |
| 10 | 3.44 | 3.25 | 2.38 | 1.7 | 1.75 | 1.56 | 0.81 | 1.06 |
| 11 | 1.13 | 1.33 | 0.75 | −0.1 | 0.63 | 1.19 | −0.31 | 1.5 |
| 12 | 2.94 | 3.33 | 3.25 | 1.95 | 0.94 | 2.5 | 2.38 | 4 |
| 13 | 2.13 | 1.29 | 1.13 | 1.15 | −0.81 | 1.19 | 2.19 | 1.5 |
| 14 | 4.5 | 4.42 | 4.63 | 4.25 | 4.25 | 4.38 | 4.25 | 4.25 |
| 15 | 3.38 | −1.29 | 0.25 | 2.95 | −1.5 | 2.94 | 3.19 | 3.69 |
| 16 | 1.56 | 1.04 | 3.63 | 3.35 | −0.56 | 3.13 | 4.75 | 3.5 |
| 17 | 2.31 | 1.96 | 1.63 | 1.85 | 1.94 | 2.06 | 2.63 | 2.44 |
| 18 | 0.44 | 0.08 | 0.63 | 0.55 | −0.13 | 0.44 | 0.13 | 0 |
| 19 | 4.5 | 2.67 | 4 | 3.4 | 0.44 | 3.56 | 3.5 | 3.38 |
| 20 | 3.56 | 1.33 | 4.25 | 1.35 | 1.44 | 4.44 | 3.63 | 3.06 |
| 21 | 3.5 | 0.67 | 2.63 | 2 | −0.19 | 3.81 | 1.69 | 5 |
| 22 | 0.75 | −0.71 | 1 | −0.05 | 0.19 | 1 | 2.06 | 0.94 |
| 23 | 1.19 | −0.71 | 0.38 | 1.4 | 2.25 | 1.38 | 1.5 | 1.13 |
| 24 | −0.25 | −0.29 | 2.75 | 3 | 0.44 | 3.25 | 2.25 | 3.75 |
| 25 | 2 | −0.42 | 0.13 | 0.9 | −0.56 | 1.44 | 1.19 | 1.63 |
| 26 | 3 | 2.88 | 2.38 | 2.8 | 3 | 2.69 | 2.94 | 3.13 |
| 27 | −1.38 | −1.5 | −0.88 | 0.4 | 0.19 | 0.88 | −1.06 | 0.06 |
| 28 | −2.75 | −3.33 | −2.88 | −2.6 | −2.88 | −2.25 | −3.56 | −3.44 |
| 29 | 3.06 | 3.83 | 4 | 1 | 2.19 | 2 | 2.88 | 3 |
| 30 | 2.13 | 1.29 | 2.13 | 1 | 1.88 | 1.44 | 2.06 | 1.44 |

## Appendix C. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.artint.2016.08.003.

## References

[1] C. Anderson, O.P. John, D. Keltner, A.M. Kring, Who attains social status? Effects of personality and physical attractiveness in social groups, J. Pers. Soc. Psychol. 81 (1) (2001) 116–132.

[2] Y. Bachrach, E. Markakis, A.D. Procaccia, J.S. Rosenschein, A. Saberi, Approximating power indices, in: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, vol. 2, AAMAS'08, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2008, pp. 943–950, http://dl.acm.org/citation.cfm?id=1402298.1402354.

[3] M. Baptista, H. Damas, J. Dimas, G. Raimundo, R. Prada, C. Martinho, P. Santos, J. Pena, A serious game based on a public goods experiment, in: 2013 International Conference on Social Computing, SocialCom, Sep. 2013, pp. 774–781.
[4] J. Bates, Virtual reality, art, and entertainment, Presence, J. Teleoper. Virtual Environ. 1 (1) (1992) 133–138.
[5] G. Boella, L. Sauro, L. van der Torre, From social power to social importance, Web Intell. Agent Syst. 5 (4) (Dec. 2007) 393–404, http://dl.acm.org/citation.cfm?id=1377783.1377786.
[6] Y. Bourse, Artificial intelligence in the Sims series, http://yoannbourse.com/ressources/docs/ens/sims-slides.pdf, Feb. 2012.
[7] S. Brainov, T. Sandholm, Power, dependence and stability in multiagent plans, in: Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Conference on Innovative Applications of Artificial Intelligence, AAAI'99/IAAI'99, American Association for Artificial Intelligence, Menlo Park, CA, USA, 1999, pp. 11–16, http://www.aaai.org/Library/AAAI/1999/aaai99-002.php.
[8] M.E. Bratman, D.J. Israel, M.E. Pollack, Plans and resource-bounded practical reasoning, Comput. Intell. 4 (3) (Sep. 1988) 349–355, http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8640.1988.tb00284.x/abstract.
[9] M. Brauer, R.Y. Bourhis, Social power, Eur. J. Soc. Psychol. 36 (4) (Jul. 2006) 601–616, http://onlinelibrary.wiley.com/doi/10.1002/ejsp.355/abstract.
[10] P. Busch, D.T. Wilson, An experimental analysis of a salesman's expert and referent bases of social power in the buyer–seller dyad, J. Mark. Res. 13 (1) (Feb. 1976) 3, http://www.jstor.org/stable/3150896.
[11] J. Campos, C. Martinho, G. Ingram, A. Vasalou, A. Paiva, My dream theatre: putting conflict on center stage, in: 8th International Conference on the Foundations of Digital Games, 2013, pp. 283–290, http://www.fdg2013.org/program/papers/paper37_campos_etal.pdf.
[12] C. Carabelea, O. Boissier, C. Castelfranchi, Using social power to enable agents to reason about being part of a group, in: M.-P. Gleizes, A. Omicini, F. Zambonelli (Eds.), Engineering Societies in the Agents World V, vol. 3451, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 166–177, http://www.springerlink.com/content/5vbaby2xy3m6aqpm/.
[13] D. Cartwright, Power: a neglected variable in social psychology, Stud. Soc. Power 6 (1959) 1–14.
[14] C. Castelfranchi, Social power: a point missed in multi-agent, DAI, and HCI, in: Y. Demazeau, J. Müller (Eds.), Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW-89, Elsevier, 1990, pp. 49–62.
[15] C. Castelfranchi, Modelling social action for AI agents, Artif. Intell. 103 (1–2) (Aug. 1998) 157–182, http://www.sciencedirect.com/science/article/pii/S0004370298000563.
[16] C. Castelfranchi, The micro–macro constitution of power, ProtoSociology 18–19 (2003) 208–268.
[17] C. Castelfranchi, InMind and OutMind Societal Order Cognition & Self-Organization: The Role of MAS, May 2013.
[18] C. Castelfranchi, M. Miceli, A. Cesta, Dependence relations among autonomous agents, ACM SIGOIS Bull. 13 (3) (1992) 14.
[19] P. Costa, R. McCrae, Revised NEO Personality Inventory (NEO PI-R) and Neo Five-Factor Inventory (NEO-FFI), Psychological Assessment Resources, Odessa, FL, 1992.
[20] R.A. Dahl, The concept of power, Behav. Sci. 2 (3) (1957) 201–215, http://onlinelibrary.wiley.com/doi/10.1002/bs.3830020303/abstract.
[21] M.H. Davis, A multidimensional approach to individual differences in empathy, JSAS Cat. Select. Doc. Psychol. 10 (85) (1980).
[22] J.D. DeLamater, Handbook of Social Psychology, Springer, New York, 2006, http://www.springer.com/social+sciences/book/978-0-387-32515-6.
[23] J. Dias, S. Mascarenhas, A. Paiva, FAtiMA modular: towards an agent architecture with a generic appraisal framework, in: T. Bosse, J. Broekens, J. Dias, J.v.d. Zwaan (Eds.), Emotion Modeling, in: Lect. Notes Comput. Sci., vol. 8750, Springer International Publishing, 2014, pp. 44–56.
[24] F. Dignum, R. Prada, G.J. Hofstede, From autistic to social agents, in: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems, AAMAS'14, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2014, pp. 1161–1164, http://dl.acm.org/citation.cfm?id=2617388.2617431.
[25] E. Doirado, C. Martinho, I mean it!: detecting user intentions to create believable behaviour for virtual agents in games, in: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, AAMAS'10, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2010, pp. 83–90, http://dl.acm.org/citation.cfm?id=1838206.1838218.
[26] M.B. Donnellan, F.L. Oswald, B.M. Baird, R.E. Lucas, The Mini-IPIP scales: tiny-yet-effective measures of the big five factors of personality, Psychol. Assess. 18 (2) (2006) 192–203, http://dx.doi.org/10.1037/1040-3590.18.2.192.
[27] C. Elliott, J. Brzezinski, Autonomous agents as synthetic characters, AI Mag. 19 (2) (Jun. 1998) 13, https://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1366.
[28] R.M. Emerson, Power-dependence relations, Am. Sociol. Rev. 27 (1) (Feb. 1962) 31–41, http://www.jstor.org/stable/2089716.
[29] T. Falbo, Multidimensional scaling of power strategies, J. Pers. Soc. Psychol. 35 (8) (1977) 537–547, http://dx.doi.org/10.1037/0022-3514.35.8.537.
[30] T. Falbo, L.A. Peplau, Power strategies in intimate relationships, J. Pers. Soc. Psychol. 38 (4) (1980) 618–628, http://dx.doi.org/10.1037/0022-3514.38.4.618.
[31] R. Falcone, C. Castelfranchi, Issues of trust and control on agent autonomy, Connect. Sci. 14 (4) (2002) 249–263, http://www.tandfonline.com/doi/abs/10.1080/0954009021000068763.
[32] A. Field, D.G.J. Hole, How to Design and Report Experiments, 1st edition, Sage Publications Ltd, London, Thousand Oaks, Calif., Feb. 2003.
[33] A.P. Field, Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock 'n' Roll, SAGE, London, 2013.
[34] S.T. Fiske, J. Berdahl, Social power, in: Social Psychology: Handbook of Basic Principles, 2nd ed., Guilford Press, New York, NY, US, 2007, pp. 678–692, http://www.guilford.com/books/Social-Psychology/Kruglanski-Higgins/9781572309180.
[35] D.R. Forsyth, Are groups bad for your health?, in: Group Dynamics, 5th edition, Wadsworth Publishing Company, Mar. 2009, p. 148.
[36] D.R. Forsyth, Group Dynamics, 5th edition, Wadsworth Publishing Company, Mar. 2009.
[37] J. French Jr., B. Raven, The bases of social power, in: Studies in Social Power, Univer. Michigan, Oxford, England, 1959, pp. 150–167.
[38] P.F. Gomes, MEEMOs: believable agents with episodic memory retrieval, Ph.D. thesis, Universidade Técnica de Lisboa – Instituto Superior Técnico, 2010.
[39] D. Hayes, H. Hexmoor, Social power as an exchangeable resource for distributed multi-agent systems, in: Proceedings of the International Symposium on Collaborative Technologies and Systems, IEEE Computer Society, Washington, DC, USA, 2006, pp. 278–281, http://dl.acm.org/citation.cfm?id=1153169.1153363.
[40] H. Hexmoor, Absolute model of autonomy and power: toward group effects, Connect. Sci. 14 (4) (2002) 323–333.
[41] R.W. Hill Jr., J. Belanich, H.C. Lane, M. Core, M. Dixon, E. Forbell, J. Kim, J. Hart, Pedagogically structured game-based training: development of the ELECT BiLAT simulation, Tech. rep., DTIC Document, 2006, http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA461575.
[42] T.R. Hinkin, C.A. Schriesheim, Development and application of new scales to measure the French and Raven (1959) bases of social power, J. Appl. Psychol. 74 (4) (1989) 561–567.
[43] A.J.I. Jones, M. Sergot, A formal characterisation of institutionalised power, Log. J. IGPL 4 (3) (Jun. 1996) 427–443, http://jigpal.oxfordjournals.org/content/4/3/427.
[44] S. Karkoulian, L. Messarra, M. Sidani, Correlates of the bases of power and the big five personality traits: an empirical investigation, J. Organ. Cult. Commun. Confl. 13 (2) (2009).
[45] D. Keltner, G.A. Van Kleef, S. Chen, M. Kraus, A reciprocal influence model of social power: emerging principles and lines of inquiry, Adv. Exp. Soc. Psychol. 40 (2008) 151–192, http://www.sciencedirect.com/science/article/pii/S0065260107000032.
[46] T.D. Kemper, Status, Power and Ritual Interaction, Ashgate, Sep. 2011.

[47] P.G. Kenny, A. Hartholt, J. Gratch, W. Swartout, D. Traum, S.C. Marsella, D. Piepol, Building interactive virtual humans for training environments, in: Interservice/Industry Training, Simulation and Education Conference, I/ITSEC, Orlando, FL, Nov. 2007, http://ict.usc.edu/pubs/Building%20Interactive%20Virtual%20Humans%20for%20Training%20Environments.pdf.
[48] M. Koslowsky, J. Schwarzwald, The power interaction model: theory, methodology, and empirical applications, in: A.Y. Lee-Chai, J.A. Bargh (Eds.), The Use and Abuse of Power: Multiple Perspectives on the Causes of Corruption, Psychology Press, New York, NY, US, 2001, pp. 195–214.
[49] D.E. Krause, Influence-based leadership as a determinant of the inclination to innovate and of innovation-related behaviors: an empirical investigation, Leadersh. Q. 15 (1) (Feb. 2004) 79–102, http://www.sciencedirect.com/science/article/pii/S1048984303001036.
[50] G. Lakoff, M. Johnson, Metaphors We Live By, University of Chicago Press, Dec. 2008.
[51] K. Lewin, Field Theory in Social Science: Selected Theoretical Papers, edited by Dorwin Cartwright, Harpers, 1951, http://psycnet.apa.org/psycinfo/1951-06769-000.
[52] F. López, Social power and norms: impact on agent behaviour, Ph.D. thesis, University Of Southampton, 2003.
[53] N. Magnenat-Thalmann, Z. Kasap, Virtual humans in serious games, in: International Conference on CyberWorlds, CW'09, Sep. 2009, pp. 71–79.
[54] S. Marsella, D. Pynadath, Modeling influence and theory of mind, in: Artificial Intelligence and the Simulation of Behavior, 2005, pp. 199–206.
[55] S. Mascarenhas, R. Prada, A. Paiva, N. Degens, G.J. Hofstede, "Can I ask you a favour?": a relational model of socio-cultural behaviour, in: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'13, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2013, pp. 1335–1336, http://dl.acm.org/citation.cfm?id=2484920.2485211.
[56] S. Mascarenhas, R. Prada, A. Paiva, G.J. Hofstede, Social importance dynamics: a model for culturally-adaptive agents, in: R. Aylett, B. Krenn, C. Pelachaud, H. Shimodaira (Eds.), Intelligent Virtual Agents, in: Lect. Notes Comput. Sci., vol. 8108, Springer, Berlin, Heidelberg, Aug. 2013, pp. 325–338.
[57] D.C. McClelland, Human Motivation, CUP Archive, 1987.
[58] S. Milgram, Obedience to Authority: An Experimental View, Harper & Row, 1974.
[59] I. Millington, J. Funge, Artificial Intelligence for Games, CRC Press, Aug. 2009.
[60] M.S. Nesler, H. Aguinis, B.M. Quigley, S.-J. Lee, J.T. Tedeschi, The development and validation of a scale measuring global social power based on French and Raven's power taxonomy, J. Appl. Soc. Psychol. 29 (4) (Apr. 1999) 750–769, http://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.1999.tb02022.x/abstract.
[61] V.L. Patel, D.R. Kaufman, T. Cohen, Cognitive Informatics in Health and Biomedicine: Case Studies on Critical Care, Complexity and Errors, Springer Science & Business Media, Nov. 2013.
[62] J. Pfeffer, Power in Organizations, Pitman, Marshfield, MA, 1981.
[63] W.E. Pollard, T.R. Mitchell, Decision theory analysis of social power, Psychol. Bull. 78 (6) (Dec. 1972) 433–446, http://www.sciencedirect.com/science/article/pii/S0033290907648138.
[64] R. Prada, A. Paiva, Teaming up humans with autonomous synthetic characters, Artif. Intell. 173 (1) (2009) 80–103.
[65] A.S. Rao, M.P. Georgeff, et al., BDI agents: from theory to practice, in: ICMAS, vol. 95, 1995, pp. 312–319, http://www.aaai.org/Papers/ICMAS/1995/ICMAS95-042.
[66] B. Raven, A power/interaction model of interpersonal influence: French and Raven thirty years later, J. Soc. Behav. Pers. 7 (2) (1992) 217–244.
[67] B. Raven, J. Schwarzwald, M. Koslowsky, Conceptualizing and measuring a power/interaction model of interpersonal influence, J. Appl. Soc. Psychol. 28 (4) (1998) 307–332.
[68] B.H. Raven, The bases of power and the power/interaction model of interpersonal influence, Anal. Soc. Issues Public Policy 8 (1) (Dec. 2008) 1–22, http://onlinelibrary.wiley.com/doi/10.1111/j.1530-2415.2008.00159.x/abstract.
[69] M.O. Riedl, A. Stern, Believable agents and intelligent scenario direction for social and cultural leadership training, in: Proceedings of the 15th Conference on Behavior Representation in Modeling and Simulation, BRIMS, May 15–18, 2006, Baltimore, MD, 2006, Paper 27.
[70] D. Rousseau, B. Hayes-Roth, A social–psychological model for synthetic actors, in: Proceedings of the Second International Conference on Autonomous Agents, AGENTS'98, ACM, New York, NY, USA, 1998, pp. 165–172, http://doi.acm.org/10.1145/280765.280795.
[71] J. Sabater, C. Sierra, Reputation and social network analysis in multi-agent systems, in: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1, AAMAS'02, ACM, New York, NY, USA, 2002, pp. 475–482, http://doi.acm.org/10.1145/544741.544854.
[72] W.H. Seto, T.Y. Ching, Y.B. Chu, W.L. Seto, Social power and motivation for the compliance of nurses and housekeeping staff with infection control policies, Am. J. Infect. Control 19 (1) (Feb. 1991) 42–44, http://www.sciencedirect.com/science/article/pii/019665539190159A.
[73] J.S. Sichman, Depint: dependence-based coalition formation in an open multi-agent scenario, J. Artif. Soc. Soc. Simul. 1 (2) (1998), http://jasss.soc.surrey.ac.uk/1/2/3.html.
[74] J.S. Sichman, R. Conte, Multi-agent dependence by dependence graphs, in: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1, AAMAS'02, ACM, New York, NY, USA, 2002, pp. 483–490, http://doi.acm.org/10.1145/544741.544855.
[75] J.S. Sichman, R. Conte, Y. Demazeau, C. Castelfranchi, A social reasoning mechanism based on dependence networks, in: Proceedings of 11th European Conference on Artificial Intelligence, 1994, pp. 188–192, http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.9217.
[76] S. Sutherland, Irrationality: The Enemy Within, vol. ix, Constable and Company, London, England, 1992.
[77] J.L. Swasy, Measuring the bases of social power, Adv. Consum. Res. 6 (1) (1979) 340–346.
[78] P.A.M. Van Lange, J.A. Joireman, How we can promote behavior that serves all of us in the future, Soc. Issues Policy Rev. 2 (1) (2008) 127–157, http://onlinelibrary.wiley.com/doi/10.1111/j.1751-2409.2008.00013.x/abstract.
[79] N. Vannini, S. Enz, M. Sapouna, D. Wolke, S. Watson, S. Woods, K. Dautenhahn, L. Hall, A. Paiva, E. Andre, R. Aylett, FearNot!: a computer-based anti-bullying-programme designed to foster peer intervention, Eur. J. Psychol. Educ. 26 (1) (2010) 21–44.
[80] R.E. Wray, J.E. Laird, A. Nuxoll, D. Stokes, A. Kerfoot, Synthetic adversaries for urban combat training, AI Mag. 26 (3) (2005) 82, http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1828.
[81] S. Younger, Leadership in small societies, J. Artif. Soc. Soc. Simul. 13 (3) (2009) 5.