Katia Sycara
Michele Gelfand
Allison Abbe   *Editors*

Models for Intercultural Collaboration and Negotiation

Springer

Models for Intercultural Collaboration and Negotiation

Advances in Group Decision and Negotiation

Volume 6

The book series, Advances in Group Decision and Negotiation — as an extension of the journal, *Group Decision and Negotiation* — is motivated by unifying approaches to group decision and negotiation processes. These processes are purposeful, adaptive and complex – cybernetic and self-organizing – and involve relation and coordination in multiplayer, multicriteria, ill-structured, evolving dynamic problems in which players (agents) both cooperate and conflict. These processes are purposeful complex adaptive systems.

Group decision and negotiation involves the whole process or flow of activities relevant to group decision and negotiation – such as, communication and information sharing, problem definition (representation) and evolution, alternative generation, social emotional interaction, coordination, leadership, and the resulting action choice.

Areas of application include intraorganizational coordination (as in local/global strategy, operations management and integrated design, production, finance, marketing and distribution – e.g., as for new products), computer supported collaborative work, labor-management negotiation, interorganizational negotiation (business, government and nonprofits), electronic negotiation and commerce, mobile technology, culture and negotiation, intercultural and international relations and negotiation, globalization, terrorism, environmental negotiation, etc.

For further volumes:
http://www.springer.com/series/5587

Katia Sycara • Michele Gelfand • Allison Abbe
Editors

Models for Intercultural Collaboration and Negotiation

Editors

Katia Sycara
School of Computer Science
Carnegie Mellon University
Pittsburgh Pennsylvania
USA

Michele Gelfand
Department of Psychology
University of Maryland
College Park Maryland
USA

Allison Abbe
U.S. Army Research Institute
    for the Behavioral & Social Sciences
Arlington Virginia
USA

Foreword

Over the last five decades the corporate world has increasingly seen the value in expanding their reach across international borders. Organizations have worked to enter new markets overseas, gain access to cheaper labor pools, and take advantage of potential synergies and new ideas through merging with similar organizations in other nations. In doing so, they have encountered, and worked to solve, the variety of problems that stem from interacting and working across national and cultural boundaries. In the last decade, the military community has similarly encountered and begun work to solve many of these problems. In 2008 the Department of the Army requested that a consortium of universities led by Dr. Michele Gelfand at the University of Maryland and Dr. Katia Sycara at Carnegie Mellon University, undertake a line of research examining critical social processes across cultural boundaries. This research has focused on understanding collaboration and negotiation in cross-cultural contexts. The consortium is composed of two research teams led respectively by Drs. Gelfand and Sycara, with each team approaching these research problems from two complementary perspectives. One team has begun from a psychological perspective building grounded theory to explain observed behavior in negotiation and collaboration activity. A separate team has worked from a computational modeling perspective to model the observed computationally and thus inform theory on negotiation and collaboration. These complementary approaches, combining a psychologically-grounded approach with computational modeling, hold great promise for making significant strides forward in our understanding of these critical social processes in cross-cultural contexts.

As the reader makes their way through this volume, they will note the tension between the reductionism necessary for computational models to capture behavior with some measure of accuracy and the contextualization necessary for fully understanding behavior in the real world. Balancing this tension is both intentional in the design of the research effort and necessary to achieving the knowledge desired. The research teams represented in these chapters have provided an excellent example of how these two fields can cooperatively – indeed collaboratively – work together to achieve more than either could in isolation.

Unlike some volumes which have an inherently retrospective tone, the intent of this book is to capture and convey the surge forward in thinking in this domain and push our collective understanding forward. The editors and authors articulate the new insights they have gained through this research enterprise and explore new challenges they have uncovered along the way. This research effort is not yet complete, and thus the discussion contained here is but a useful pause to reflect and report the knowledge gained to date. It is also a point of departure for empirically exploring the immediate questions raised. Further these ideas and questions may inspire the many post-doctoral fellows, graduate and undergraduate students to continue this line of inquiry in the future.

This type of multidisciplinary effort holds promise, and indeed may be necessary, as the most important scientific questions facing the behavioral sciences more often are found at the boundaries between scientific disciplines. It is in this challenging, often messy environment that some of the most significant scientific achievements of our time will be made. As such, the editors and authors – the principal researchers undertaking this effort – are to be congratulated for their willingness to undertake such an effort and persistence in handling the many challenges they have faced along the way. Multidisciplinary research is not easy. However, when executed well, as it has been here, it can be extraordinarily productive.

Chief of Basic Research   Gerald F. Goodwin
U.S. Army Research Institute

Contents

A Formal Definition of Culture

Aliaksandr Birukou, Enrico Blanzieri, Paolo Giorgini, and Fausto Giunchiglia

Abstract Globalization makes culture no more bound to a geographical area, race or religion as was previously considered in anthropology. With the advent of Web 2.0 it becomes appropriate to speak about the culture of online communities in general, without specific ties to country or nation. Multi-national companies, software developers, scientists need to take into account cultural differences when delivering products to people. The first step in dealing with culture consists in defining and representing culture of the targeted community. AI literature addressed issues of sociality, collaboration, and coordination in agent societies, but did not target the problem of defining and representing culture of a community.

This chapter presents a formal definition of culture of a set of agents. The proposed definition generalizes existing definitions of culture and it is operational in the sense that it can be applied for characterization and comparison of culture(s) existing in various communities. The main focus of this chapter is on the first version of the formalism that does not introduce states. However, representing a snapshot of the culture in a certain moment is the first step towards a more complex formalism that includes states (the work on the latter is ongoing and we present only preliminary definitions here).

Keywords Culture • Formal definition • Agents • Communities • Web 2.0 • Traits • Actions • Artificial intelligence

A. Birukou • E. Blanzieri (✉) • P. Giorgini • F. Giunchiglia
DISI, University of Trento, via Sommarive 14, Povo, Trento, TN 38123, Italy
e-mail: birukou@disi.unitn.it; blanzieri@disi.unitn.it; pgiorgio@disi.unitn.it; fausto@disi.unitn.it

1 Introduction

The advent of Web 2.0 lead to an explosive growth in the number of applications targeted at communities, e.g. applications supporting social navigation, collaborative editing, bookmarking and tagging. In such applications, culture is no more bound to a geographical area or a religion, as it is usually studied in anthropology. It becomes appropriate to speak about the culture of online communities and such communities in general cannot be characterized in terms of race, religion, or country. Rephrasing Axelrod (1997), electronic communications allows us to develop patterns of interaction that are chosen rather than imposed by geography. Specific applications such as search engines or e-bookshops and the ways of using them become part of the culture of people. For instance, using Norton Commander file manager or preparing documents in the MS DOS 6.0 operating system, nowadays would be considered unusual to the same extent as lighting one's house with torches. Moreover, in some scenarios we can speak about societies of pure artificial agents, such as web services or programs and their specific culture, e.g. the standards implemented or the set of functionalities used. Human traders and trading agents operating on the same markets together use the same rules and develop common practices which can be referred to as culture that exists in a mixed society composed of different types of agents. All this shows that grasping and representing culture becomes an important problem in computer science. Applications should be developed consistently with the culture of the target community and the notion of culture would provide support for building such applications.

In this chapter we provide a formal definition of culture. Our goal is not to provide a formalism or a reasoning framework *per se*, but, rather, to give an operational definition of culture that can be used for characterizing and describing culture in different scenarios. In particular, we address the problems of development of applications according to the community culture and of characterizing culture of existing communities. We present and formalize a definition of the notion of culture of a set of agents at a moment in time. We define culture as a set of traits that are shared by the set, where traits are characteristics of human societies that are potentially transmitted by non-genetic means and can be owned by an agent (modified from Mulder, Nunn, & Towner, 2006). The requirement of traits being potentially transmitted is needed because transmission is a way of spreading traits, and, consequently, culture, and without transmission it is hard to achieve sharing. The sharing of such traits by the set is required for two reasons: (1) to go from the set of personal traits of an individual to the culture of the set of agents, and (2) to filter out characteristics which only pertain to the set of agents as a whole, but not to individuals. An example of latter traits is birth rate.

This chapter consolidates and slightly updates the work previously published as a technical report (Birukou, Blanzieri, Giorgini, & Giunchiglia, 2009a) and a workshop paper (Birukou, Blanzieri, Giorgini, & Giunchiglia, 2009b). We start in Sect. 2 with the review of the use of concept of culture in the literature. Section 3 presents the first version of the formalism that does not introduce states and does not

consider spreading of culture. However, representing a snapshot of the culture of a set of agents in a certain moment is the first step towards a more complex formalism that includes states and models transmission. Section 4 defines some measures on top of the formalism, and an example, illustrating our approach, is presented in Sect. 5. A preliminary version of the formalism with the states is there presented in Sect. 6 We discusses related work and limitations of the approach in Sect. 7, and conclude in Sect. 8.

2 Defining Culture

Culture is a slippery and ubiquitous concept. Initially, culture was associated with the notion of civilization tout-court. At the end of the 1930s Margaret Mead put in contrast "culture" with "a culture". "Culture means the whole complex of traditional behavior which has been developed by the human race and is successively learned by each generation" (Mead, 2002 cited in Brumann, 1999). However, specificity of the notion of culture with respect to a given human society was needed in order to study other societies. So the same citation goes on as: "A culture is less precise. It can mean the forms of traditional behavior which are characteristic of a given society, or of a group of societies, or of a certain race, or of certain area, or of a certain period of time" (cited in Brumann, 1999). As a consequence, in the anthropological literature culture has been introduced as the concept denoting the object of study of cultural anthropology. Other definitions were proposed and they largely vary. However, they seem to converge to the notion that culture is learned (Axelrod, 1997), it is associated with groups of people and its content includes a wide range of phenomena including norms, values, shared meanings, and patterned ways of behaving (Boyd & Richerson, 1988; Bozeman, 1992; Kroeber & Kluckhohn, 1952; McFate, 2005; Mead & Métraux, 1953; O'Reilly, 1989). In anthropological literature the usefulness of the notion of culture as a scientific tool has been attacked giving rise to the so-called "writing against culture movement" (see Brumann, 1999 for a reaction against it). The culture as defined in anthropology usually refers to societies defined in national or ethnic terms, however, the concept of culture has been recently used for describing knowledge and behavior of other groups like in the concepts of corporate culture or organizational culture (Harrison & Carroll, 2006; O'Reilly, 1989; Schein, 1996). Moreover, globalization has brought about the problem of interaction of cultures. On the one hand, such interaction leads to blurring boundaries between cultures, while on the other hand it leads to the increasing need of cultural-aware managers and professionals. Recent anthropology textbook definitions take into account the shift in meaning as, for example, in the definition by Peoples and Bailey:

> *Culture is the socially transmitted knowledge and behavior shared by some group of people*
> (Bailey & Peoples, 1998, p. 23 cited in Brumann, 1999).

Earlier authors define culture in the following ways (cited in Brumann, 1999):

– Culture ... refers ... to learned, accumulated experience. A culture ... refers
  to those socially transmitted patterns for behavior characteristic of a particular
  social group (Keesing, 1981, p. 68).
– Culture, or civilization, ... is that complex whole which includes knowledge,
  belief, art, law, morals, custom, and any other capabilities and habits acquired by
  man as a member of society (Tylor, 1871, p. 1).
– The culture of any society consists of the sum total of ideas, conditioned
  emotional responses, and patterns of habitual behavior which the members of
  that society have acquired through instruction or imitation and which they share
  to a greater or less degree (Linton, 1936).
– A culture is the total socially acquired life-way or life-style of a group of people.
  It consists of the patterned, repetitive ways of thinking, feeling, and acting that
  are characteristic of the members of a particular society or segment of a society
  (Harris, 1975).

As we can see, definitions agree on the fact that culture consists of something
that is shared and/or learned by a group of people, but the content of the culture
varies in different definitions. Similarly to Axelrod (1997), we see the content of
the culture as a set of traits,[1] which can refer to behavior, knowledge facts, ideas,
beliefs, norms, etc.

3   A Formal Definition of Culture. The Version
Without States

Consistently with the AI literature, we define an *agent* as a "physical or virtual
entity that can act, perceive its environment (in a partial way) and communicate
with others, is autonomous and has skills to achieve its goals and tendencies"
(Ferber, 1999). An agent can represent an individual or a collective entity such
as an organization, and can have different *cultural traits*, which are characteristics
of human societies that are potentially transmitted by non-genetic means and can
be owned by an agent. The requirement "can be owned by", which we add to the
definition by Mulder et al. (2006), means that it is possible for an agent to have a
cultural trait. Behavior, beliefs, knowledge are particular kinds of cultural traits.

Let us consider the set of agents Ag and the set of traits $\mathcal{T}$. Given an agent $a \in Ag$
we denote its set of cultural traits with $T_a = \{\tau_i\} \subseteq \mathcal{T}$ and we use the predicate
$has(a, \tau)$ to represent the fact that the agent a has a trait $\tau \in T_a$. In the following,
we call the set of traits of an individual *the culture of an individual*.

---

[1]Traits are further grouped in *features* in Axelrod's formulation, i.e. each feature can take value
from a set of specific traits.

Table 1 The set of traits $\mathcal{T}$ in Example 1

Trait type	Traits
Knowledge	*Dante_Alighieri_wrote_Divine_Comedy(DA), cappuccino_is_coffee(CI),*
	latte_macchiato_is_coffee(LM), Meiji_era_was_in_1868_1912(ME)
Behavior	*Eating_with_sticks(ES), eating_with_fork(EF)*
Norms, rules	*Never_put_mayonnaise_on_pizza(NP),*
	never_open_umbrella_inside_building(NO)
Beliefs	*Christianity(Chr), Buddhism(Bud)*

Table 2 Traits of agents in Example 1

Set	Traits
$T_{Charlie}$	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee,*
	cappuccino_is_coffee, eating_with_sticks, eating_with_fork,
	never_put_mayonnaise_on_pizza, Buddhism
T_{Pedro}	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee,*
	cappuccino_is_coffee, eating_with_fork, Christianity
T_{Toru}	*Meiji_era_was_in_1868_1912, cappuccino_is_coffee, eating_with_sticks, Buddhism*
T_{Maria}	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee,*
	cappuccino_is_coffee, eating_with_sticks, eating_with_fork, Christianity
T_{Andrea}	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee,*
	cappuccino_is_coffee, eating_with_fork, Christianity

Example 1. Let *Ag* in our example be a set of people: Charlie, Pedro, Maria, and Andrea are European citizens, and Toru is from Japan. Let $\mathcal{T}$ be a set of traits of different types, as shown in Table 1. For each trait, we also put its abbreviation (used in the figures in this section) in parentheses.

Table 2 lists the set of traits $\mathcal{T}$, and the sets of traits of the specific agents of $Ag = \{$ *Charlie, Pedro, Toru, Maria, Andrea* $\}$.

We can write *has(Maria, Dante_Alighieri_wrote_Divine_Comedy)*, or *has(Charlie, cappuccino_is_coffee)*, but not *has(Andrea, eating_with_sticks)*. We will use this example as a running example through the section. ⊘

Note that we introduce types of traits and use them in the example only for convenience. One might propose a different classification of traits, e.g. putting *eating_with_fork* as a norm. We believe that there is no single classification and this suggests that our approach of dealing with generic traits rather than with specific types of cultural content provides certain advantages.

Definition 1 (sharing). For each pair of agents $a_i, a_j \in Ag$ and for each trait $\tau \in \mathcal{T}$, a_i and a_j share the trait τ iff they both have such a trait:

$$has(a_i, \tau) \wedge has(a_j, \tau) \leftrightarrow sharing(a_i, a_j, \tau).$$

Property 1. Sharing is transitive:

$$sharing(a_i, a_j, \tau) \wedge sharing(a_j, a_k, \tau) \rightarrow sharing(a_i, a_k, \tau).$$
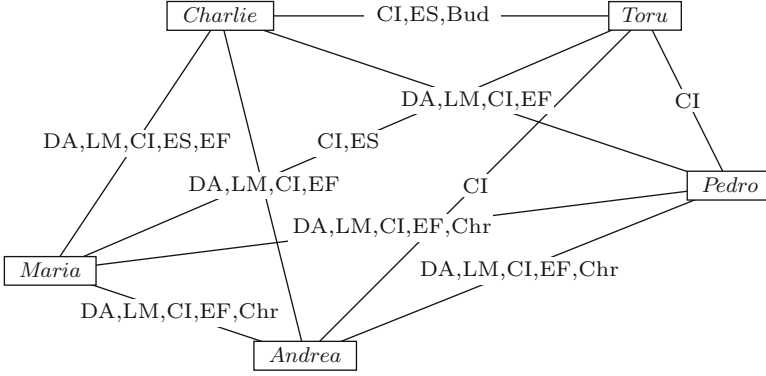
Fig. 1 The graph showing for which agents and traits the predicate *sharing* holds in Example 1. The nodes are agents and labels on each edge denote traits that are shared by the pair of agents connected by the edge. For instance, the edge between *Toru* and *Andrea* labeled *CI* means that *sharing*(*Andrea*, *Toru*, *cappuccino_is_coffee*). The traits are abbreviated as in Table 1: *Dante_Alighieri_wrote_Divine_Comedy* is abbreviated as *DA*, *latte_macchiato_is_coffee* as *LM*, *cappuccino_is_coffee* as *CI*, *eating_with_sticks* as *ES*, *eating_with_fork* as *EF*, *Christianity* as *Chr*, *Buddhism* as *Bud*

Fig. 2 The graph that shows for which agents the *sharing* predicate holds for the *cappuccino_is_coffee*(*CI*) trait in Example 1
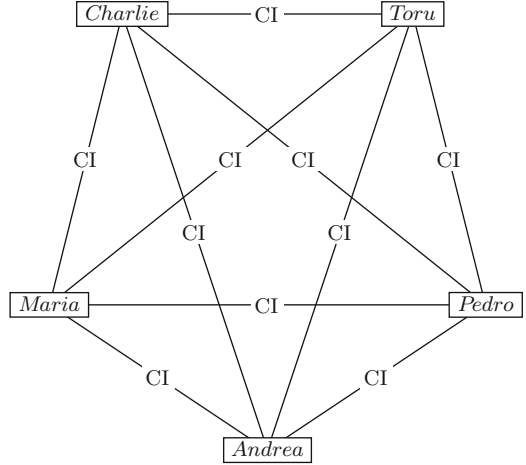
Example 1 (continued). In the example, we can write *sharing* (*Toru*, *Maria*, *eating_with_sticks*), or *sharing*(*Pedro*, *Andrea*, *cappuccino_is_coffee*), etc. To avoid giving the complete list of tuples for which *sharing* holds, we represent them as a graph where nodes are agents and labels on each edge denote traits that are shared by the pair of agents connected by the edge, see Fig. 1.

We can show a restriction of *sharing* to specific traits, as in Fig. 2, which shows how one trait, *cappuccino_is_coffee*, is shared by the set of agents.   ⊘

Given a set of agents $G \subseteq Ag$ and a set of traits $T_G \subseteq \mathcal{T}$ we define the notions of *weak sharing* and *strong sharing*.

Definition 2 (weak sharing). A set of traits T_G is weakly shared by a set of agents G iff for each trait $\tau \in T_G$ there exists a pair of agents $a_i, a_j \in G, \ a_i \neq a_j$ that share τ.

Definition 3 (strong sharing). A set of traits T_G is strongly shared by a set of agents G iff each trait $\tau \in T_G$ is shared by all pairs of agents $a_i, a_j \in G, \ a_i \neq a_j$.

Example 1 (continued). Let us consider the set of agents $G = \{$ *Charlie*, *Toru*, *Maria*, *Andrea*, *Pedro* $\}$ and two sets of traits: $T_G = \{$ *cappuccino_is_coffee*, *eating_with_sticks*, *Dante_Alighieri_wrote_Divine_Comedy* $\}, T'_G = \{$ *cappuccino_is_coffee* $\}$. Using the *sharing* predicate represented in Fig. 1, we can see that T'_G is strongly shared by G because the *cappuccino_is_coffee* trait is shared by each pair of agents. T_G contains three traits that are shared by at least one pair of agents: e.g., *cappuccino_is_coffee*, *eating_with_sticks* shared by *Toru* and *Charlie*, *Dante_Alighieri_wrote_Divine_Comedy* shared by *Charlie* and *Andrea*. So, T_G is weakly shared by G.   ⊘

Property 2. Strong sharing implies weak sharing.

Proof. Strong sharing of a set of traits T_G by a set of agents G means that for each $\tau \in T_G$ all pairs of agents $a_i, a_j \in G$ share τ. Thus, the condition for weak sharing, i.e. existence of one pair of agents $a_i, a_j \in G, \ a_i \neq a_j$ that share τ is fulfilled. □

Given a set of agents $G \subseteq Ag$ such that $|G| \geq 2$, we introduce the notion of culture of G.

Definition 4 (weak culture of a set of agents). A non-empty set of traits $T_G \subseteq \mathcal{T}$ is a weak culture of G iff

- the set T_G is weakly shared by G,
- for each agent $a \in G$ there exists a trait $\tau \in T_G$ such that $has(a, \tau)$.

In other words, Definition 4 says that for a set of agents, a culture is defined as a set of traits weakly shared by agents, and each agent has at least one trait in the culture.

Definition 5 (strong culture of a set of agents). If T_G in Definition 4 is also strongly shared then it is *a strong culture of a set of agents*.

In the following if we refer to "a culture of a set of agents", we mean "a weak culture of a set of agents".

Example 1 (continued). We first give examples of sets of traits that are not culture because one of the conditions for being culture is not satisfied and then give an example of weak and strong cultures.

Let us consider the set $G = \{$ *Pedro*, *Maria* $\}$, and the set $T_G = \{$ *eating_with_sticks* $\}$. T_G is not a culture of G, because T_G is not weakly shared by G.

Let us add *Toru* in the set. Now, the set $T_{G'} = \{\,eating\_with\_sticks\,\}$ is not a culture of $G' = \{\,Pedro, Maria, Toru\,\}$ because *Pedro* does not have the *eating_with_sticks* trait.

On the other hand, the set $T'_{G'} = \{\,Dante\_Alighieri\_wrote\_Divine\_Comedy,$ *eating_with_sticks* $\}$ is a weak culture of G' since the traits *eating_with_sticks* and *Dante_Alighieri_wrote_Divine_Comedy* are weakly shared by G' and each agent has at least one trait in $T'_{G'}$ (*Toru* and *Maria* have *eating_with_sticks*, *Pedro* has *Dante_Alighieri_wrote_Divine_Comedy*).

Taking $G'' = \{\,Toru, Maria\,\}$, $T_{G''} = \{\,eating\_with\_sticks\,\}$ is a strong culture, because it is strongly shared by the set, and each agent has the *eating_with_sticks* trait.   ⊘

Property 3. Given a set of agents $G \subseteq Ag$ and T_G, a culture of G, it is possible to find a non-empty set $G_0 \subseteq G$ and a non-empty set T_{G_0} such that T_{G_0} is a strong culture of G_0.

Proof. If $|G| = 2$ then all traits that are weakly shared are also strongly shared and $T_{G_0} = T_G$ is a strong culture of $G_0 = G$. Otherwise, let us consider $G'_0 = \{\,a_1, a_2\,\}$, where a_1 and a_2 are two agents of G such that $T_{a_1} \cap T_{a_2} \cap T_G \neq \emptyset$ (agents share at least one trait from the culture). The existence of such a pair of agents is guaranteed, because T_G is weakly shared, so for every $\tau \in T_G$ there are at least two agents that share it. Since both a_1 and a_2 have each trait from $T_{G_0} = T_{a_1} \cap T_{a_2} \cap T_G$, it is strongly shared and the second condition required for a set to be a strong culture is fulfilled.   □

4   Measures for the Comparison of Cultures

In this section, we present some measures for characterizing a culture of a set of agents in different socio-cultural settings and for comparing cultures of different sets. This list is not exhaustive, rather, it contains some initial measures, and further extension of this list is a subject of future research.

4.1   Measuring Culture as a Snapshot

4.1.1   Culture

Let us start from simple measures such as *presence of a specific trait in a culture*. We use an indicator function $I_{has}(\tau, T)$ to say that the trait τ is present in the culture T:

$$I_{has}(\tau, T) = \begin{cases} 1, & \text{if } \tau \in T \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Another example of a simple measure of a culture could be *the number of traits in the culture*, defined as $|T|$, i.e. the dimension of the set of traits T.

4.1.2 Culture of a Group

A culture of a group is a product of the individuals belonging to the group. However, different groups can share cultures to some extent. To measure such degree of sharing we adapt the notion of cultural homogeneity introduced by Carley, in 1991. Culture in that paper is defined as the distribution of information (ideas, beliefs, concepts, technical knowledge, etc.) across population. In our settings, given a set of agents G and a culture T_G of G, the cultural homogeneity is measured by the percentage of possible dyadically shared traits that actually are shared. A trait τ is shared by a dyad if $sharing(a_i, a_j, \tau)$. The number of possible dyadically shared traits is $\binom{N}{2} \times K$, where N is the number of agents in the set, $|G|$; $K = |T_G|$ is the number of traits in the culture T_G. Thus, *cultural homogeneity* is measured as

$$CH(G, T_G) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{K} I_{sharing}(a_i, a_j, \tau_k)}{\binom{N}{2} \times K} \times 100\%. \tag{2}$$

In this formula, $G = \{a_i\}$, $1 \leq i \leq N$, $T_G = \{\tau_k\}$, $1 \leq k \leq K$, and the indicator function $I_{sharing}$ is defined as follows:

$$I_{sharing}(a_i, a_j, \tau_k) = \begin{cases} 1, & \text{if } sharing(a_i, a_j, \tau_k) \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to note that the cultural homogeneity takes into account only traits present in the culture, and it does not matter what traits agents of G have besides those contained in the culture T_G. To take the traits that are not a part of culture into account, we introduce the notion of group homogeneity. To do this, we need to consider the set of all traits of the group $\bar{T}_G = \cup_{i=1}^{N} a_i$, $\bar{K} = |\bar{T}_G|$. Thus, *group homogeneity* of the group G is measured as

$$GH(G) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{\bar{K}} I_{sharing}(a_i, a_j, \tau_k)}{\binom{N}{2} \times \bar{K}} \times 100\%, \tag{3}$$

where τ_k, $1 \leq k \leq \bar{K}$ are from the set $\bar{T}_G$ and the other terms are defined in Eq. 2.

4.1.3   A Culture of an Individual and a Culture of a Group

To compare a culture of an individual a and a culture of a group G we introduce the following measures:

- *Common culture (culture overlap)* is the set of traits that is present in both cultures: $CC(T_a, T_G) = T_a \cap T_G$.
- *Culture similarity* is the degree to which two cultures are similar, i.e. how much they have in common: $CS(T_a, T_G) = \frac{|T_a \cap T_G|}{|T_a \cup T_G|} \times 100\%$.
- *Culture fit* is the degree to which one culture fits the other culture: $CF(T_a, T_G) = \frac{|T_a \cap T_G|}{|T_G|} \times 100\%$. Note that this measure is not symmetric.

Note that it is possible to extend the notion of culture similarity further if we assume there is a domain-specific function for calculating similarity between traits, i.e. for each pair of traits τ_1, τ_2 we know the value of $sim(\tau_1, \tau_2)$. Culture similarity can then be defined as $CC(T_a, T_G) = \frac{\sum_{i=1}^{|T_a|} \sum_{j=1}^{|T_G|} sim(\tau_i^a, \tau_j^G)}{|T_a| \times |T_G|} \times 100\%$ This will allow for considering the degree of similarity between different traits, e.g., specifying that trait *eating_with_sticks* is more similar to *eating_with_fork* than to *telling_DA*.

4.1.4   A Culture of a Group and a Culture of Another Group

In order to compare cultures of two sets of agents we can straightforwardly replace the culture of an individual with a culture of another group in the formulas above, thus introducing the following measures:

- Common culture (culture overlap) is the set of traits that is present in both cultures: $CC(T_{G_1}, T_{G_2}) = T_{G_1} \cap T_{G_2}$.
- Culture similarity is the degree to which two cultures are similar, i.e. how much they have in common: $CS(T_{G_1}, T_{G_2}) = \frac{|T_{G_1} \cap T_{G_2}|}{|T_{G_1} \cup T_{G_2}|} \times 100\%$.
- Culture fit is the degree to which one culture fits the other culture: $CF(T_{G_1}, T_{G_2}) = \frac{|T_{G_1} \cap T_{G_2}|}{|T_{G_2}|} \times 100\%$.
  Note that this measure is not symmetric.

Table 3 summarizes the presented measures.

4.2   Example

Let us see how the described measures apply to Example 1 from Sect. 3, summarized in Table 2. Considering a set of agents $G = \{$ *Charlie*, *Toru*, *Andrea*, *Maria*, *Pedro* $\}$, and a culture $T_G = \{$ *Dante_Alighieri_wrote_Divine_Comedy(DA)*, *eating_with_sticks(ES)* $\}$:

Table 3 Measures of culture as a snapshot

Measure	Meaning
$I_{has}(\tau, T)$	Shows if the trait τ is present in the culture T
$|T|$	The number of elements in the culture T
$CH(G, T_G)$	Cultural homogeneity of G, i.e. how widely the culture T_G is shared within the group G
$GH(G)$	Group homogeneity, i.e. how similar are the sets of traits of agents of G
$CC(T_a, T_G)$	Common culture, i.e. the set of traits contained in the culture of an agent a
$CC(T_{G_1}, T_{G_2})$	(a group G_1) and in the culture T_G (T_{G_1})
$CS(T_a, T_G)$	Culture similarity, i.e. how much two cultures have in common
$CS(T_{G_1}, T_{G_2})$	
$CF(T_a, T_G)$	Culture fit, i.e. the degree to which the culture of a (G_1) fits the culture T_G (T_{G_2})
$CF(T_{G_1}, T_{G_2})$	

– $I_{has}(eating\_with\_sticks, T_G) = 1$,
– $I_{has}(eating\_with\_fork, T_G) = 0$,
– $|T_G| = 2$.

To calculate the cultural homogeneity of G we need to calculate the number of traits in the culture T_G: $K = 2$ and the number of agents in the set G: $N = 5$. With these parameters, $CH(G, T_G)$ is calculated as follows:

$$CH(G, T_G) = \frac{\sum_{i=1}^{5} \sum_{j=i+1}^{5} \sum_{k=1}^{2} I_{sharing}(a_i, a_j, \tau_k)}{\binom{5}{2} \times 2} \times 100\%$$

$$= \frac{\sum_{i=1}^{5} \sum_{j=i+1}^{5} (I_{sharing}(a_i, a_j, DA) + I_{sharing}(a_i, a_j, ES))}{10 \times 2} \times 100\%$$

$$= \frac{6 + 3}{20} \times 100\% = 45\%.$$

Proceeding with calculations we get:

– $GH(G) = \frac{41}{120} \times 100\% = 34.17\%$,
– $CC(Pedro, T_G) = \{Dante\_Alighieri\_wrote\_Divine\_Comedy\}$,
– $CS(Pedro, T_G) = 0.125$,
– $CF(Pedro, T_G) = 0.5$,
– $CF(T_G, Pedro) = 0.143$.

5 A Case Study

In this section, we provide a case study that shows how the material presented in this chapter can be applied in the Web 2.0 domain. We first describe the scenario and then show how it can be addressed with our approach.

5.1 Scenario Description

Let us consider activities related to bibliography management in CiteULike.org, a free online service to organize someone's collection of academic papers. Users of CiteULike are mainly scientists and there are groups dedicated to specific interests. The site allows people to add papers in their personal collections or to the collections of the groups users belong to and to tag those papers. It is also possible to search for the papers using keywords or browse the papers with a specific tag.

Let us suppose that Michael, a user of CiteULike, has some papers about recommendation systems in his bibliography and has tagged them as shown in Table 4.[2] He discovers that there are groups on CiteULike and that there are at least three groups that seem relevant to his research interests: GroupA, GroupB, and GroupC. In the group bibliography, each group has a list of papers tagged as shown in Table 4. Michael would like to join some group, but he does not have much time to read group feeds, so he would like to choose only one group. How does he decide which group fits more with his interests? The bibliography of a group contains several hundred of items, looking through them will take some time.

Let us assume that all tags are from the same taxonomy and there are no syntactical (e.g., tags *recommendation system*, *recommender systems*, *RS* are replaced with a single tag) and semantical (e.g., tags like *recommendation system*, *adaptive system* correspond to very same concepts in all bibliographies) inconsistencies in the names of papers and tags. Thus, we can represent a group or a user as a set of tags and a set of papers in their bibliography and calculate the degree of the fit between a user and a group as similarity between their sets of tags and papers. Moreover, we can see which papers are common for all three groups, creating for Michael a list of papers to read.

5.2 Applying Our Approach

In our formalism, the users and groups are agents that are represented as a set of traits, which are papers and tags. For each agent, its culture is the set of traits as follows:

---

[2]Of course, we present a simplified example here, real users and groups on CiteULike have much more papers in their bibliographies.

Table 4 Users and groups in CiteULike.org

PaperID	Paper	Tags
Michael		
PolyLens	PolyLens: a recommender system for groups of users	Recommendation, collaborative filtering
TrustInRS	Trust in recommender systems	Trust, recommendation
GroupLens	GroupLens: An Open Architecture for Collaborative Filtering of Netnews	Collaborative filtering, grouplens
RefWeb	Referral Web: Combining Social Networks and Collaborative Filtering	Collaborative filtering, trust
TrustCF	Trust-Aware Collaborative Filtering for Recommender Systems	Trust, recommendation
GroupA		
EComRec	E-Commerce Recommendation Applications	Collaborative filtering, e-commerce, recommender
TechLens	Enhancing digital libraries with TechLens+	Recommender, academic reference
GetToKnow	Getting to know you: learning new user preferences in recommender systems	Collaborative filtering, recommender
GroupLens	Group Lens: an open architecture for collaborative filtering of netnews	Collaborative filtering, recommender
PolyLens	PolyLens: a recommender system for groups of users	Recommendation, collaborative filtering
GroupB		
TechLens	Enhancing Digital Libraries with TechLens+	Collaborative filtering, content based filtering, papers, recommender systems
Citations	On the Recommending of Citations for Research Papers	Citations, collaborative filtering, personalization, recommender systems
Scouts	Scouts, promoters, and connectors: the roles of ratings in nearest-neighbor collaborative filtering	Recommender systems, recommendation, collaborative filtering
EComRec	E-Commerce Recommendation Applications	Collaborative filtering, e-commerce, recommender
ContRec	A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation	Collaborative filtering, concept extraction, concept map, recommender

Table 4 (continued)

PaperID	Paper	Tags
GroupC		
GroupLens	Group Lens: an open architecture for collaborative filtering of netnews	Collaborative filtering, recommender, recommendation
VirtCom	Recommending and evaluating choices in a virtual community of use	Collaborative filtering, recommender
TagCF	Tag-aware recommender systems by fusion of collaborative filtering algorithms	Tagging, recommender, collaborative filtering
TrustInRS	Trust in recommender systems	Trust, recommender, collaborative filtering
RefWeb	Referral Web: Combining Social Networks and Collaborative Filtering	Collaborative filtering, social network

Michael.papers={ *PolyLens*, *TrustInRS*, *GroupLens*, *RefWeb*, *TrustCF* }
Michael.tags={ *recommendation*, *collaborative filtering*, *trust*, *grouplens* }
GroupA.papers={ *EComRec*, *TechLens*, *GetToKnow*, *GroupLens*, *PolyLens* }
GroupA.tags={ *collaborative filtering*, *recommendation*, *academic reference*, *recommender*, *e-commerce* }
GroupB.papers={ *TechLens*, *Citations*, *Scouts*, *EComRec*, *ContRec* }
GroupB.tags={ *collaborative filtering*, *content based filtering*, *papers*, *citations*, *recommender systems*, *personalization*, *recommendation*, *e-commerce*, *recommender*, *concept extraction*, *concept map* }
GroupC.papers={ *GroupLens*, *VirtCom*, *TagCF*, *TrustInRS*, *RefWeb* }
GroupC.tags={ *collaborative filtering*, *recommender*, *recommendation*, *tagging*, *trust*, *social network* }

Let us select one of the metrics from Sect. 4, say culture similarity, for determining how close are two cultures. Since the number of distinct papers in *Michael*'s and *GroupA* bibliographies is eight, the number of common papers is two, the number of distinct tags is seven and the number of common tags is two, the similarity between *Michael* and *GroupA*, *CS*(*Michael*, *GroupA*) is equal to $0.5 \cdot \frac{2}{8} + 0.5 \cdot \frac{2}{7} = 0.268$. The similarity between *Michael* and *GroupB* is $0.5 \cdot \frac{0}{10} + 0.5 \cdot \frac{2}{13} = 0.077$, while the similarity between *Michael* and *GroupC* is $0.5 \cdot \frac{3}{7} + 0.5 \cdot \frac{3}{7} = 0.429$. From this simple exercise we can conclude that *Michael*'s research interests, as represented by his bibliography, are closer to *GroupC*. The program realizing such algorithm in real CiteULike.org settings, i.e. with hundreds of groups with thousands of papers, would solve the above-mentioned problem of choosing which community to join.

Let us further illustrate how our formalism can be applied to these data. Let us consider each group as an agent and see which traits are shared by the set of agents { *GroupA*, *GroupB*, *GroupC* }. Papers *EComRec*, *TechLens*, *GroupLens* and

tags *recommender*, *e-commerce*, *recommendation*, *collaborative filtering* are weakly shared by the set and therefore are a culture of the set. Moreover, while there are no strongly shared papers, tags *collaborative filtering*, *recommender*, *recommendation* are strongly shared and therefore are a strong culture of the set.

5.3 Discussion

In the case study we calculated the degree of culture similarity between Michael and different groups, and computed a culture of a set of CiteULike groups. Further extending this example, we might take into account not only artifacts such as papers or tags, but also behaviors of users, such as tagging some paper with a specific tag. For instance, using information about authors of the papers and citations, it is possible to consider behaviors such as self-citation and to see if there are communities whose members follow this practice more than an average author. Using information about the publication date and the date of posting the publication in someone's library it is possible to consider behaviors such as "tagging paper before its publication" and see which communities have the practice of dissemination of drafts of the papers.

6 A Formal Definition of Culture. The Version with States

This section presents a preliminary version of the definition of culture with states. Traits and agents are defined as in Sect. 3.

To model changes in the set of traits of an agent and consequently, changes in culture, we use the notion of state. We assume that the world can be in different states and the set of traits of the same agent can be different in different states.

Let us consider the set of agents Ag, the set of traits $\mathcal{T}$, and the set of states S. Given an agent $a \in Ag$ and a state $s \in S$, we denote the set of cultural traits of the agent a in the state s with $T_a(s) = \{ \tau_i \} \subseteq \mathcal{T}$ and we use the predicate $has(a, \tau_i, s)$ to represent the fact that the agent a has a trait $\tau_i \in T_a(s)$ in the state s. In the following, we call the set of traits of an individual *the culture of an individual*.

Example 2. Let us consider a set of people and model them as agents with a set of traits and a behavior related to transmission, *telling_DA* (telling that Dante Alighieri wrote "The Divine Comedy"). Let Ag in our example be a set of people: Charlie, Pedro, Maria, and Andrea are European citizens, and Toru is from Japan. Let $\mathcal{T}$ be a set of traits of different types.

Table 5 lists the sets of traits of the specific agents of $Ag = \{$ *Charlie*, *Pedro*, *Toru*, *Maria*, *Andrea* $\}$ in the state s_1. Changes with respect to Table 2 are highlighted in bold. We can write $has(Maria, Dante\_Alighieri\_wrote\_Divine\_Comedy, s_1)$, or $has(Charlie, cappuccino\_is\_coffee, s_1)$, but not $has(Andrea, eating\_with\_sticks, s_1)$. We will use this example as a running example. $\oslash$

Table 5 Traits of agents in Example 2

Set	Traits
$T_{Charlie}(s_1)$	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee*, **telling_DA**, *cappuccino_is_coffee, eating_with_sticks, eating_with_fork, never_put_mayonnaise_on_pizza, Buddhism*
$T_{Pedro}(s_1)$	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee, cappuccino_is_coffee, eating_with_fork, Christianity*
$T_{Toru}(s_1)$	*Meiji_era_was_in_1868_1912, cappuccino_is_coffee, eating_with_sticks, Buddhism*
$T_{Maria}(s_1)$	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee, cappuccino_is_coffee, eating_with_sticks, eating_with_fork, Christianity*
$T_{Andrea}(s_1)$	*Dante_Alighieri_wrote_Divine_Comedy, latte_macchiato_is_coffee, cappuccino_is_coffee, eating_with_fork, Christianity*

We distinguish behavior as a particular kind of traits and assume that performing a behavior by an agent changes the state of the world. We define the set of all behaviors $\mathcal{B} \subseteq \mathcal{T}$ and the function *perform* in $Ag \times \mathcal{B} \times S \rightarrow S$. The intended meaning of this function is that an agent, which has some behavior in some state, performs this behavior in this state and the state of the world changes to another state. More specifically, $s_v = perform(a, \tau, s_u)$ means that $has(a, \tau, s_u)$ and the agent *a* performed a behavior τ in the state s_u and the resulting state is s_v. The fact that $has(a, \tau, s_u)$ does not imply that the agent *a* is able to perform the behavior τ in the state s_u, because some preconditions for performing the behavior may be not fulfilled in the state s_u. Note that since traits are not innate, by assuming $\mathcal{B} \subseteq \mathcal{T}$ we do not include innate behaviors, such as blinking when air is puffed in someone's eye.

At this point we would like to discuss the distinction between *action* and *behavior*. In AI literature, an action is an atomic piece of activity, while behavior is perceived as something more complex, and can include several actions. Therefore, our notion of performing a behavior can really be decomposed into performing several actions. However, we decided not to introduce explicit relations between actions and behaviors. Moreover, the absence of such clear dependency in AI literature suggests that these relations are hard or even impossible to formalize. Instead, we assume that behavior can represent an atomic action or a more complex activity depending on the level of modeling granularity. We can vary granularity of behaviors depending on the problem in hand and on the domain. For instance, in Example 2, when someone needs to know whether agents are working, it is possible to consider behaviors *working* and *playing*, or, even, *working* and *not_working*. However, if someone would like to have a closer look at leisure activities of the group, it is necessary to introduce finer granularity of the *playing* behavior, e.g. by considering *playing_basketball* and *playing_chess* behaviors.

We assume that the states are ordered, we define recursively the order "is before" and the corresponding predicate *is_before*(s_u, s_v) and *is_after*(s_v, s_u) in the following way:

Definition 6 (is_before). $is\_before(s_u, s_v) \leftrightarrow \exists a \in Ag, \ \tau \in \mathcal{B}, \ s \in S$ such that $s = perform(a, \tau, s_u) \wedge (s = s_v \vee is\_before(s, s_v))$.

Definition 7 (is_after). $is\_after(s_v, s_u) \leftrightarrow is\_before(s_u, s_v)$

We assume that in each state s_v, the previous state s_u is uniquely defined, while the next state depends on the action an agent performs in s_v. From Definition 6 we can derive the following property:

Property 4. For all agents $a \in Ag$, for all behaviors $\tau \in \mathcal{B}$ and for all states $s_u, s_v \in S$

$$s_v = perform(a, \tau, s_u) \rightarrow is\_before(s_u, s_v)$$

Definition 8 (sharing). For each pair of agents $a_i, a_j \in Ag$, for each trait $\tau \in \mathcal{T}$, and for each state $s \in S$, a_i and a_j share the trait τ in the state s iff they both have such a trait in s:

$$has(a_i, \tau, s) \wedge has(a_j, \tau, s) \leftrightarrow sharing(a_i, a_j, \tau, s). \tag{4}$$

We also assume that agents do not lose traits when the state of the world changes, as the following axiom says:

Axiom 1. *For all agents $a \in Ag$, traits $\tau \in \mathcal{T}$, and states $s \in S$:*

$$has(a, \tau, s) \rightarrow \forall s_v : is\_after(s_v, s) \ has(a, \tau, s_v).$$

Example 2 (continued). In the example, we can write *sharing(Toru,Maria, eating_with_sticks, s_1)*, or *sharing(Pedro, Andrea, cappuccino_is_coffee, s_1)*, etc. Traits that are shared by a pair of agents in the state s_1 are as in Fig. 1. ⊘

Let us assume that if an agent a_i has a trait τ, the trait τ can be transmitted to another agent a_j before some state s and we use the predicate *transmitted(a_i, a_j, τ, s)* to represent this. We represent *transmitted(a_i, a_j, τ, s)* in a graph by a directed edge from a_i to a_j labeled τ.

Axiom 2. *For each pair of agents $a_i, a_j \in Ag$, $a_i \neq a_j$, for each trait $\tau \in \mathcal{T}$, and for each state $s \in S$ the fact that the trait τ has been transmitted from a_i to a_j before the state s implies that exists some state $s_u \in S$ such that a_i has τ in the state s_u, a_j does not have τ in the state s_u and an agent a_k performing a behavior τ_m in the state s_u imply that in the resulting state s_v the agent a_j has τ:*

$$transmitted(a_i, a_j, \tau, s) \rightarrow (\exists s_u \in S, \ is\_before(s_u, s) \wedge has(a_i, \tau, s_u) \wedge$$

$$\neg has(a_j, \tau, s_u) \wedge (s_v = perform(a_k, \tau, s_u)) \rightarrow has(a_j, \tau, s_v))$$

From our assumption that traits are not innate, it follows that traits are acquired by agents, and the goal of the transmitted predicate is to show the way an agent acquired a trait. Therefore, we assume that in the initial state agents have no traits and the way they acquire traits is represented using the transmitted predicate.
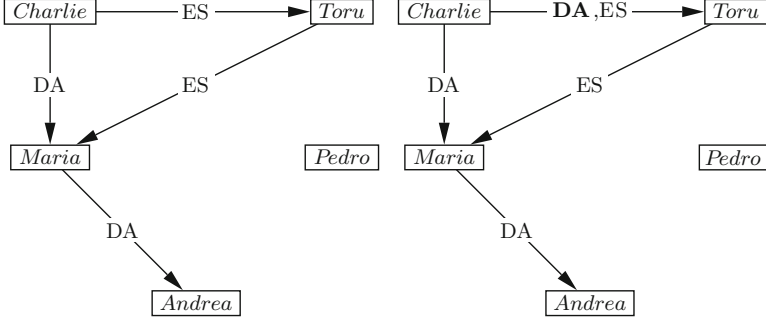
Fig. 3 The graph that shows for which agents the *transmitted* predicate holds in the state s_1 (*left*) and s_2 (*right*) in Example 2. Changes with respect to the state s_1 are in *bold*

We should note that the trait τ is not shared by a_i and a_j in the state s_u, while it is shared by a_i and a_j in the state s_v, and in the state s, as shown by the following property:

Property 5. For all pairs of agents $a_i, a_j \in Ag$, for all traits $\tau \in \mathcal{T}$, and for all states $s_v \in S$

$$sharing(a_i, a_j, \tau, s_v) \rightarrow (\forall s : is\_after(s, s_v) \ sharing(a_i, a_j, \tau, s))$$

From Axiom 2 it also follows that the transmitted predicate holds for all subsequent states after s_v.

Property 6. For all pairs of agents $a_i, a_j \in Ag$, for all traits $\tau \in \mathcal{T}$, and for all states $s_v \in S$

$$transmitted(a_i, a_j, \tau, s_v) \rightarrow (\forall s : is\_after(s, s_v) \ transmitted(a_i, a_j, \tau, s))$$

Example 2 (continued). Figure 3 shows the graph representing the *transmitted* predicate in state s_1 in our example. The traits *Dante_Alighieri_wrote_Divine_Comedy* and *eating_with_sticks* have been transmitted. On the contrary, the traits *cappuccino_is_coffee* and *never_put_mayonnaise_on_pizza* have not been transmitted (the latter trait is not even shared by any pair of agents). In particular, the *Dante_Alighieri_wrote_Divine_Comedy* trait has been transmitted from *Charlie* to *Maria*, and from *Maria* to *Andrea*. Also, the *eating_with_sticks* trait has been transmitted from *Charlie* to *Toru* and from *Toru* to *Maria*. We can write *transmitted*(*Charlie*, *Maria*, *Dante_Alighieri_wrote_Divine_Comedy*, s_1).

Let us assume that in the state s_1 *Charlie* tells *Toru* that Dante Alighieri wrote the Divine Comedy and *Toru* memorizes this piece of knowledge. This corresponds to $s_2 = perform(Charlie, telling\_DA, s_1)$. The *transmitted* predicate in the state s_2 is as depicted in the left part of Fig. 3 and *transmitted* in the state s_2 is as depicted in the right part of Fig. 3. The difference in the *transmitted* predicates
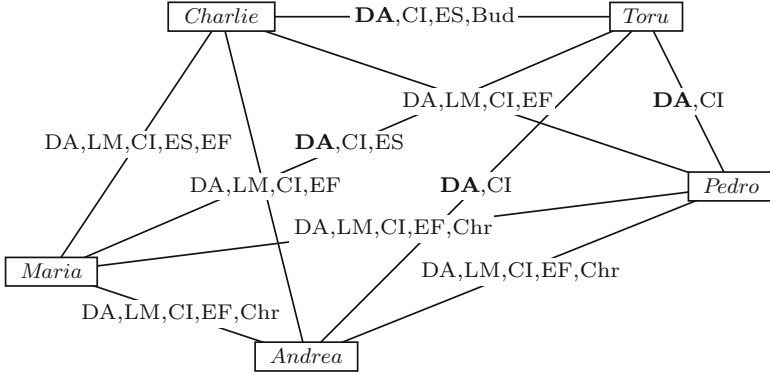
Fig. 4 The graph that shows for which agents the *sharing* predicate holds in the state s_2 in Example 2. Changes with respect to the state s_1 are in *bold*

in these two states is that the *Dante_Alighieri_wrote_Divine_Comedy* trait has been transmitted from *Charlie* to *Toru* and the corresponding edge is added, namely *transmitted*(*Charlie*, *Toru*, *Dante_Alighieri_wrote_Divine_Comedy*, s_2). In the state s_2 the following change in the set of traits for Toru occurs: $T_{Toru}(s_2) = \{Meiji\_era\_was\_in\_1868\_1912$, *Dante_Alighieri_wrote_Divine_Comedy*, *cappuccino_is_coffee*, *eating_with_sticks*, *Buddhism*$\}$.

Obviously, the transmission has an impact on sharing and the *sharing* predicate in the state s_2 is as depicted in Fig. 4, with the edges between *Toru* and *Charlie*, *Maria*, *Andrea*, *Pedro* added.   ⊘

Given a set of agents $G \subseteq Ag$ and a set of traits $T_G \subseteq \mathcal{T}$ we define the notions of *weak sharing* and *strong sharing*.

Definition 9 (weak sharing). A set of traits T_G is weakly shared by a set of agents G in a state s iff for each trait $\tau \in T_G$ there exists a pair of agents $a_i, a_j \in G$, $a_i \neq a_j$ that share τ in the state s.

Definition 10 (strong sharing). A set of traits T_G is strongly shared by a set of agents G in a state s iff each trait $\tau \in T_G$ is shared by all pairs of agents $a_i, a_j \in G$ in s.

In other words, the set of traits is weakly (strongly) shared if it is a subset of the union (intersection) of traits shared by pairs of agents of G in the state s.

Example 2 (continued). Let us consider the set of agents $G = \{Charlie, Toru, Maria, Andrea, Pedro\}$. Analyzing the *sharing* predicate in the state s_1 we can see that only the *cappuccino_is_coffee* trait is shared by each pair of agents in the state s_1, so $T_G = \{cappuccino\_is\_coffee\}$ is strongly shared by G in the state s_1. There are three traits that are shared by at least one pair of agents in the state s_1: *cappuccino_is_coffee*, *eating_with_sticks* shared, e.g, by *Toru* and *Charlie*, and *Dante_Alighieri_wrote_Divine_Comedy* shared, e.g., by *Charlie* and *Andrea*.

So, the set $T'_G = \{$*Dante_Alighieri_wrote_Divine_Comedy*, *cappuccino_is_coffee*, *eating_with_sticks*$\}$ and all non-empty subsets of this set are weakly shared by the set G in the state s_1. Analogously, the set $T''_G = \{$*eating_with_sticks*, *Dante_Alighieri_wrote_Divine_Comedy*, *cappuccino_is_coffee*$\}$ is weakly shared by G in the state s_3, and the set $T'''_G = \{$*cappuccino_is_coffee*, *Dante_Alighieri_wrote _Divine_Comedy*$\}$ is strongly shared by the set G in the state s_3. $\oslash$

Property 7. Strong sharing implies weak sharing.

Given a set of agents $G \subseteq Ag$ such that $|G| \geq 2$, and a *transmitted* predicate we introduce the notion of culture of G.

Definition 11 (weak culture of a set of agents). A non-empty set of traits $T_G \subseteq \mathcal{T}$ is a weak culture of G in a state s iff

– the set T_G is weakly shared by G in the state s,
– for each agent $a \in G$ in the state s there exists a trait $\tau \in T_G$ such that $has(a, \tau, s)$.

From the assumption that traits are not innate, as we discussed, it follows that traits are acquired by agents, as represented by the transmitted predicate. Therefore, we can formulate the following axiom, telling that all traits in culture are transmitted.

Axiom 3. *For each trait $\tau \in T_G$ there exists an agent $a \in Ag$ that transmitted τ to another agent $a_j \in G$ before the state s, i.e. transmitted(a, a_j, τ, s).*

From Definition 11 and Axiom 3 it follows that all the traits in the culture are transmitted, shared, and each agent has at least one trait from the culture. Please, note that since the traits are transmitted not necessarily within the set, the transmitted predicate does not imply sharing between the agents of G.

Definition 12 (strong culture of a set of agents). If T_G in Definition 11 is also strongly shared in the state s then it is *a strong culture of the set of agents G in the state s.*

In the following if we refer to "a culture of a set of agents", we mean "a weak culture of a set of agents".

Example 2 (continued). Considering $G = \{$*Toru, Andrea*$\}$ in the state s_3, $T_G = \{$*Dante_Alighieri_wrote_Divine_Comedy*, *cappuccino_is_coffee*$\}$ is strongly shared by the set G in the state s_3.

Although the *Dante_Alighieri_wrote_Divine_Comedy* trait has been transmitted both to *Toru* and *Andrea* from outside (from *Charlie* and *Maria*, respectively), it is strongly shared by the agents of G. Since in the state s_3 each agent in G has the trait *Dante_Alighieri_wrote_Divine_Comedy*, $T'_G = \{$*Dante_Alighieri_wrote_Divine_Comedy*$\}$ is a culture of G in the state s_3. It is easy to see that T'_G is not a culture of G in the states s_1 and s_2 because *Toru* does not have this trait in those states. $\oslash$

The following proposition outlines some restrictions on how culture can change between states, namely it shows that culture is monotonic.

Proposition 1 (monotonicity of culture). *If a non-empty set of traits T_G is a culture of a set of agents G in a state s_v, then T_G is a culture of G also in any state s after s_v.*

In real world, the traits of a culture can be lost for two reasons: (1) agents can lose traits, (2) agents can die, move to another group, etc. As we stated in Axiom 1, in our model, agents do not lose traits. However, our model, and the proposition about monotonicity of culture support the case when agents disappear from the group.

Definition 13 (union culture of a group). A non-empty set of traits T_G^{union} is the union culture of a set of agents G in the state s iff T_G^{union} is the union of all cultures T_G of G in the state s.

In other words, the union culture of a set of agents in some state is the union of all possible cultures of the set in this state. Since it is the union of all cultures, it is not possible to add any trait to T_G^{union} and still obtain a culture of G. In the following, we refer to the union culture of a set of agents as "the culture of a set".

Definition 14 (evolution of culture). A sequence of sets of traits $\{ T_G^{(1)}, \ldots, T_G^{(i)} \}$ is an evolution of culture of G iff:

- exists a sequence of states $\{ s_1, \ldots, s_i \}$, such that $T_G^{(k)}$ is a culture of G in the state s_k for all k, $1 \le k \le i$,
- for each k, $1 \le k \le i - 1$ holds $is\_after(s_{k+1}, s_k)$.

In other words, a sequence of sets of traits is an evolution of culture if each set of traits in the sequence is a culture of G in some state and the states are ordered in the same way as the sets of traits. We denote evolution of culture as $\{ T_G \}$.

7   Related Work and Discussion

There are a number of approaches that are related to the development of applications for communities: social navigation (Dieberger, 1997), community-based personalization (Freyne, Smyth, Coyle, Balfe, & Briggs, 2004), open source software development to mention a few. However, existing approaches develop applications that support the culture pre-defined by the designer rather than actual culture of the community where the application is being deployed. For instance, in 1998 NetCaptor browser, followed by Opera in 2000 and Mozilla Firefox in 2001, introduced tabbed browsing, thus opening a new way to browsing in the Internet. As time passed, more and more people became familiar with tabbed browsing and this lead other browsers, e.g. Internet Explorer, to introduce tabbed browsing. We argue that tabbed browsing can be considered an element of culture, rarely found in

the culture of Internet users of 90s, but very common nowadays,[3] and this change
forced Microsoft to change the pre-defined culture of supporting only one-page-
one-window browsing to the actual culture of tabbed browsing.

The AI literature on agents addresses the issues related to sociality, such
as social action (Castelfranchi, 1998), social co-ordination architectures and
social interaction strategies for decentralized co-ordination in multi-agent systems
Ossowski (1999), social laws and conventions in multi-agent environments Shoham
and Tennenholtz (1995, 1997), and social roles (Masolo et al., 2004). However, the
issue of sociality alone does not help either to understand what differentiates one
set of agents from another or to grasp what are the specificities of the behavior of
agents of a specific society. Although in two different agent societies agents can be
able to communicate with each other and perform other social actions, these two
societies can be very different from each other. We claim that the concept of culture
can be used to describe and compare sets of agents. Some research papers use the
notion of culture in the context of agents, see e.g. Axelrod (1997), Carley (1991),
and Epstein and Axtell (1996), other provide a model for the comparison of cultures
(Hofstede, 2001). However, none of the previous research works provides a formal
definition of culture that could be readily adopted for building applications for
communities and applied for the characterization and the comparison of culture.

Carley (1991) considers culture as the distribution of information (ideas, beliefs,
concepts, symbols, technical knowledge, etc.) across the population and proposes a
model for knowledge transfer based on interactions. In that model, the probability
of an interaction between two agents is based on the principle of homophily, i.e. the
greater the amount of knowledge they share the more probable the interaction is.
During an interaction, agents exchange facts, so after the interaction one of the
agents might know more than before the interaction. The knowledge transfer in
these settings can be seen as a particular kind of culture spread. This work is
further extended in the Construct project (Hirshman, Carley, & Kowalchuck, 2007a,
2007b). For instance, one of the recent applications of Construct studies the effects
of different methods of information diffusion on spreading beliefs and knowledge
about illegal tax schemes in different American cities (Hirshman, Birukou, Martin,
Bigrigg, & Carley, 2008). With respect to the definition of culture we propose in
this paper, that model of information diffusion is complementary, because it models
transmission of elements of culture (e.g., beliefs, knowledge) in a society.

Axelrod (1997) considers culture as a list of features or dimensions of culture.
Each feature represents an individual attribute that is subject to social influence and
can have different values called traits. Two individuals have the same culture if they
have the same traits for all features. Similarly to the work by Carley, feature of an
agent can change its value during an interaction and the probability of interaction is
based on the homophily.

The notion of trait we use in our formalism is similar to the notion of feature used
by Axelrod, specifically, each feature can take value from a set of specific traits.

---

[3]For instance, Google Chrome browser includes tabbed browsing from the first release.

Traits in our formalism also includes ideas, beliefs and technical knowledge used as culture elements by Carley. Both theories by Carley and by Axelrod are based on the assumption that culture changes as a result of an interaction. Thus, in our terms, interaction in that sense can be considered as a particular kind of transmission: there are two agents participating, it takes place in some specific state and it leads to the appearance of some cultural element in one of the agents.

Epstein and Axtell (1996) study the emergence of the group rules from local ones defined at an agent's level in an artificial society of simple agents living and consuming sugar in an artificial environment called "Sugarscape". The authors consider a culture of the society as a string of binary cultural attributes and model cultural transmission both on horizontal (between agents) and vertical (through generations) levels using simple rules. However, they do not provide any formal definition of culture since the main focus of the book is on the emergence of group rules from the local ones.

According to O'Reilly (1989), the culture of an organization is considered as strong if wide consensus exists about the content and participants believe in the importance of the content. They also formulate this as a [not necessarily big] set of values that are widely shared and strongly held. This is similar to the notion of strong culture, i.e. culture shared by all pairs of agents in a group, we consider in our formalism.

Balzer and Tuomela (2003) study social practices and the dynamics of their maintenance in groups. They define social practices as recurrent collective activities based on collective intentions. The paper focuses on informal, non-normative practices, such as playing soccer on Sundays, going to sauna on Saturday afternoon, shaking hands, sharing a ride to work. They also note that the maintenance (change, preservation, renewal) depend on the success of a practice. The main contribution of the paper is a mathematical model for the description of social practices and their maintenance in groups.

Our model of culture is not limited to social practices. Moreover, it allows for inclusion of normative practices as well. However, as a consequence, the model of Balzer and Tuomela allows for a richer description of informal social practices. For instance, our model does not permit expressing intentions, but allows operating on manifestations of activities without going into details of underlying intentions. While authors show that success of a social practice is important for its adoption, for our model it is irrelevant whether a trait is successful in some sense. Our model just captures the fact that the trait is a part of culture, no matter how it occurred. The model presented by Balzer and Tuomela is defined for groups and then goes to the individual level, thereby implementing top-down approach. In our model of culture, we start from a set of traits of an individual, consider transmission as an important means of spreading culture, and then go to the culture of a group. Thus, we implement bottom-up approach. Balzer and Tuomela, while requiring sharing of a social practice within a group, and noting the importance of transmission for spreading practice, include transmission into the model only to a certain extent, namely, considering imitation as an example of transmission. Our model of culture

allows for different types of transmission as long as there is a predicate that helps to distinguish occurred transmissions.

Hofstede (2001) treats culture as "[...] the collective programming of the mind that distinguishes the members of one group or category of people from another", proposes a model of culture and applies it for studying and comparing cultures of IBM workers in more than 50 countries. The model includes the following five independent dimensions of national culture differences: *power distance*, which is related to the different solutions to the basic problem of human inequality; *uncertainty avoidance*, which is related to the level of stress in a society in the face of an unknown future; *individualism* versus *collectivism*, which is related to the integration of individuals into primary groups; *masculinity* versus *femininity*, which is related to the division of emotional roles between men and women; and *long-term* versus *short-term* orientation, which is related to the choice of focus for people's efforts: the future or the present. Values in Hofstede's terms refer to "a broad tendency to prefer certain states over others" and are similar to attitudes and beliefs, which are just particular kind of traits in our formalism. Dimensions, similarly to Axelrod's features, take values from the set of traits. Thus, comparing with our work, the model developed by Hofstede has a different focus - it aims at comparing cultures of groups of people over several pre-defined dimensions of values, while our model supports comparison over arbitrary sets of traits. The dimensions in Hofstede's model are meant to be independent, while our formalism does not address the issue of dependency of traits, so they can be dependent on each other. In this line of thoughts, an interesting application of our model could be comparison of dependency of traits across groups, i.e. if presence of a trait or traits leads to the presence of another trait(s) for one group and to the presence of third trait(s) for another group.

The definition of culture presented here allows for representation and comparison of different cultures. However, in order to compare traits, one first needs to identify the traits of individuals. On the one hand, deducing traits from manifested behaviors of agents is not a trivial task in general. On the other hand, in specific domains this might be much easier, consider, for instance, deducing traits of users from logs of a web service, website, or an application. For instance, it would be possible to see that a group of users of a text editor always turn off the autocorrect feature and turn it off automatically in new versions of the editor prepared for this group. Taking the issue of the observability of traits into account, we see social software and Web 2.0 systems as one of the potential application domains for our model.

8 Conclusions

We have defined the notion of culture of a set of agents and we have shown that our definition can be used for comparison of communities. This definition addresses existing gaps in AI literature that deals with issues of sociality, cooperation, and negotiation, but remains oblivious to the notion of culture. The provided formalism

is a first step towards an integral approach for representing, comparing, analysing, and transferring culture of communities or group of agents. We are currently working on the extension of the formalism with the notion of culture evolution to study culture changes in Web 2.0 communities.

References

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *The Journal of Conflict Resolution, 41*(2), 203–226.

Bailey, G., & Peoples, J. (1998). *Introduction to cultural anthropology*. Belmont, CA: Wadsworth Publishing.

Balzer, W., & Tuomela, R. (2003). Collective intentions and the maintenance of social practices. *Autonomous Agents and Multi-Agent Systems, 6*(1), 7–33.

Birukou, A., Blanzieri, E., Giorgini, P., & Giunchiglia, F. (2009a). *A formal definition of culture of a set of agents. The version without states* (Technical Report DISI-09-070). Trento, Italy: University of Trento.

Birukou, A., Blanzieri, E., Giorgini, P., & Giunchiglia, F. (2009b). A formal definition of culture. In *Proceedings of the workshop on modeling intercultural collaboration and negotiation (MICON) at IJCAI'09*. http://www.cs.cmu.edu/~softagents/papers/IJCAI-09-MICON-Proceedings-final.pdf

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Bozeman, A. B. (1992). Strategic intelligence and statecraft: Selected essays. Washington, DC: Brassey's.

Brumann, C. (1999). Writing for culture: Why a successful concept should not be discarded [and comments and reply]. *Current Anthropology, 40*, S1–S27.

Carley, K. (1991). A theory of group stability. *American Sociological Review, 56*(3), 331–354.

Castelfranchi, C. (1998). Modelling social action for AI agents. textitArtificial Intelligence, 103 (1–2), 157–182.

Dieberger, A. (1997). Supporting social navigation on the world wide web. *International Journal of Human-Computer Studies, 46*, 805–825.

Epstein, J. M., & Axtell, R. L. (1996). *Growing artificial societies: Social science from the bottom up*. Cambridge, MA: The MIT Press.

Ferber, J. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence*. Harlow, UK: Addison-Wesley Professional.

Freyne, J., Smyth, B., Coyle, M., Balfe, E., & Briggs, P. (2004). Further experiments on collaborative ranking in community-based web search. *Artificial Intelligence Review, 21*(3–4), 229–252.

Harris, M. (1975). *Culture, people, nature: An introduction to general anthropology*. New York: Thomas Y. Crowell.

Harrison, R. J., & Carroll, G. R. (2006). *Culture and demography in organizations*. Princeton, NJ: Princeton University Press.

Hirshman, B. R., Carley, K. M., & Kowalchuck, M. J. (2007a). *Specifying agents in construct* (Technical Report CMU-ISRI-07-107). Pittsburgh, CA: Carnegie Mellon University.

Hirshman, B. R., Carley, K. M., & Kowalchuck, M. J. (2007b). *Loading networks in construct* (Technical Report CMU-ISRI-07-116). Pittsburgh, CA: Carnegie Mellon University.

Hirshman, B. R., Birukou, A., Martin, M. K., Bigrigg, M. W., & Carley, K. M. (2008). *The impact of educational interventions on real & stylized cities* (Technical Report CMU-ISR-08-114). Pittsburgh, CA: Carnegie Mellon University.

Hofstede, D. G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Keesing, R. M. (1981). *Cultural anthropology: A contemporary perspective*. New York: Holt, Rinehart and Winston.

Kroeber, A. L., Kluckhohn, C. (1952). Culture: A critical review of concepts and definitions. *Papers of the Peabody Museum, 47*, 643–656.

Linton, R. (1936). *The study of man: An introduction*. New York: D. Appleton-Century.

Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., et al. (2004). Social roles and their descriptions. In *Principles of knowledge representation and reasoning (KR2004)*. 267–277, AAAI Press, Palo Alto, California. http://www.informatik.uni-trier.de/~ley/db/conf/kr/kr2004.html#MasoloVBCFGG04

McFate, M. (2005). The military utility of understanding adversary culture. *Joint Force Quarterly, 38. 23*, 42–48. http://www.au.af.mil/au/awc/awcgate/jfq/1038.pdf

Mead, M. (2002). *Cooperation and competition among primitive peoples*. New Brunswick, NJ: Transaction Publishers.

Mead, M., & Métraux, R., (eds.) (1953). *The study of culture at a distance*. Chicago: University of Chicago Press.

Mulder, M. B., Nunn, C. L., & Towner, M. C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology, 15*(2), 52–64.

O'Reilly, C. (1998). Corporations, culture, and commitment: Motivation and social control in organizations. *California Management Review, 31*(4), 9–25.

Ossowski, S. (1999). *Co-ordination in artificial agent societies, social structure and its implications for autonomous problem-solving agents: Vol. 1535. Lecture notes in computer science*. Heidelberg, Germany: Springer.

Schein, E. H. (1996). *Organizational culture and leadership (Jossey-Bass Psychology Series)* (2nd ed.). San Francisco: Jossey-Bass.

Shoham, Y., & Tennenholtz, M. (1995). On social laws for artificial agent societies: Off-line design. *Artificial Intelligence, 73*(1–2), 231–252.

Shoham, Y., & Tennenholtz, M. (1997). On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence, 94*(1–2), 139–166.

Tylor, E. B. (1871) *Primitive culture*. London: J. Murray.

Toward a Contextualized Cultural Framework

Maritza R. Salazar (✉), Marissa L. Shuffler, Wendy L. Bedwell, and Eduardo Salas

Abstract This chapter provides a theoretical framework that emphasizes the importance of examining how (1) internal values, attitudes and beliefs, and (2) features of the environment shape psychological processes and behaviors within and across cultures. We define and describe how internal and external sources influence human behavior across societies, both independently and in combination. The framework presented can be used to guide theoretical and empirical work that aims to better understand the relationship between values, attitudes, beliefs and aspects of the external environment on determining human behavior. Additionally, the approach used to develop our framework may not only be useful to social scientists, but to scholars who use techniques such as computational modeling and scenario-based media studies to describe or understand responses of individuals, with varying cultural value, beliefs, and attitudes, who are embedded within particular physical settings and social situations.

M.R. Salazar (✉)
Claremont Graduate University, School of Behavioral and Organizational Science, Claremont, CA, USA
e-mail: maritza.salazar@cgu.edu

M.L. Shuffler • W.L. Bedwell • E. Salas
Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA
mshuffler@ist.ucf.edu; wbedwell@ist.ucf.edu; esalas@ist.ucf.edu

1 Introduction

Over the last several decades, cross-cultural researchers have explored how culture impacts a variety of processes that influence individual, team, and organizational outcomes (Aycan, Kanungo, & Sinha, 1999; Erez & Earley, 1987). As military, corporate, and not-for-profit organizations are increasingly operating across a variety of contexts around the globe, the insights from this research are critical for managing differences in culture and for operating in a variety of local, regional, and national contexts. Scholars suggest that nations differ in aspects beyond cultural values and emphasize the importance of understanding all of the relevant sources of influence that could potentially produce differences in psychological processes and behaviors across cultures (Busenitz, Gomez, & Spencer, 2000; Erez & Earley, 1993). In particular, better understanding of the factors that shape cross-cultural behavior across a variety of situational and social context is necessary (Matsumoto & Hee Yoo, 2006; Tsui, Nifadkar, & Ou, 2007).

Motivated by Hofstede's seminal work on culture's consequences (Hofstede, 1980a, 1980b), researchers have tended to rely on values to explain differences across communities around the world (Kirkman, Lowe, & Gibson, 2006). Cultural values consist of one's judgment about what is important in life and tend to guide individual assessments and actions within one's social environment (Hofstede, 1984; McClelland, 1961; Parsons & Shils, 1951; Rokeach, 1973). These values are believed to be shared among a group of people and transmitted from generation to generation through social learning and observation (Bandura, 1986). Studies have shown that values influence cross-cultural differences on rather than in a variety of processes and outcomes, such as cognitive attributes (Morris & Peng, 1994), motivation (Erez, 1997), negotiation styles (Gelfand & Dyer, 2000), and human resource management practices (Aycan et al., 2000).

Although the values approach to understanding culture has made quite a contribution to cross-cultural research, the notion that values are the only predictor of cultural differences is largely contested. Scholars who draw upon systems, subjective, and structuralist perspectives—which tend to focus on external behavioral influences such as norms, social networks, and other aspects of the social structure—advocate that cross-cultural differences in behavior are also influenced by the contextual environment (Gelfand, Nishii, & Raver, 2006; LeVine, 1984; Morris, Podolny, & Ariel, 2000; Triandis, 1972). Theoretically, these perspectives refute the notion that cross-cultural differences can be solely explained by beliefs, attitudes, or perceptions that reside in an individual's head (Earley & Mosakowski, 2000; Gabrenya, 1999; Morris et al., 2000). Empirical findings from a variety of studies further support this view as findings indicate that cultural differences are not always fully explained by values alone (Ip & Bond, 1995; Leung, Bond, & Schwartz, 1995). In sum, although the values approach has contributed to a greater understanding of cross-cultural issues, it is critical that our theoretical and empirical work extend beyond the values approach to include the role of external influences on behavior across cultures.

Table 1 Internal sources of influence

(A) Values
Individualism vs. collectivism
Locus of control
Masculinity/femininity
Temporal orientation
Power distance
(B) Affect
Emotional expressivity
Control
Intensity
(C) Motivation
Self enhancement/self criticism
Uncertainty reduction

This chapter offers a contribution to future studies of cultural influence by developing and presenting a framework based on a contextualized approach to understanding factors that affect psychological processes and behaviors across cultures. From this perspective, the value approach is viewed as limited because internal psychological processes can be and are fostered, primed, and shaped by a variety of aspects in the cultural context. A contextualized approach relies upon a degree of contingency between psychological and behavioral responses and influences of the environmental context, such as the social (religion and family structures), political (government and law), geographic (rural vs. urban, location, terrain characteristics), and communication (language) context. Empirically, support for this approach is evident in research that illustrates differential influences of the same situation on outcomes in different cultures, such as self-esteem and connectedness, suggesting that the external context where interaction takes place matters (Kitayama, Markus, Matsumoto, & Norasakkunkit, 1997; Morling, Kitayama, & Miyamoto, 2002).

The specific contribution of this chapter is that we provide a framework that emphasizes the importance of examining how (1) internal values, attitudes and beliefs, and (2) features of the environment shape psychological processes and behaviors across cultures, both independently and together. We explain what internal and external sources of influence consist of in Tables 1 and 2 and the resulting framework can be used to guide theoretical and empirical work that aims to better understand the relationship between values, attitudes, beliefs and aspects of the external environment in regards to human behavior. Additionally, the approach used to develop our framework may not only be useful to social scientists, but to scholars who use techniques such as computational modeling and scenario-based media studies to describe or understand responses of individuals situated in particular physical settings and social situations.

Table 2 External sources of
influence

(A) Social context
Family structure
Religion
Tightness/looseness of social norms
(B) Geography
Rural vs. urban
Climate
Latitude/longitude
(C) Language
Dialect
(D) Political context
Political system
Legal system

2 Theoretical Background

To develop our framework, it is first necessary to better understand how culture
has been defined in previous theoretical efforts. As mentioned above, different
disciplines have primarily focused on culture either as residing in one's head in
the form of values, attitudes, and beliefs (e.g., Hofestede, 1980a, 1980b, 1983,
1984; Kluckhohn & Strodtbeck, 1961; Parsons & Shils, 1951; Schwartz, 1992) or
as a system of meaning manifest in the social environment as practices, rituals,
norms, and customs (e.g., Geertz, 1973; Schweder & LeVine, 1984). These values,
attitudes, beliefs, and systems of meaning are locally defined and shaped by the
social environment in which they are embedded (Geertz, 1973). As a consequence,
these differences often distinguish communities around the globe from one another.

These differences are most often salient when culturally diverse members or
distinct cultural communities interact with one another (Wenger, 1998). Klein
(2004) suggests that these differences across collectives arise from the shared
experiences of people within them. Similar experiences with regard to (a) how
individuals are raised, (b) available resources, and (c) ecological and social pressures
all shape the formation of shared culture. She suggests that people from the
same group "…share a 'lens' for making sense of the world" (p. 256). The
cultural lens model implicitly assumes that the collective will view external cues
in a constant way, inhibiting intra-collective variation in cross-cultural behavior.
Similarly, sociologists suggest that knowledge about how to interact and relate
to others is also guided by shared understanding developed through collective
engagement in material practices and symbolic reconstruction (Bruner, 1990; Shore,
1996). As a result, a shared meaning system is thought to emerge among people who
belong to the same community or collective.

To understand cultural differences in terms of values, several taxonomies have
been created over the past few decades. Cultural value-based taxonomies, such as
Hofstede's Cultural Taxonomy (1980a) and other models of culture (Kluckhohn

& Strodtbeck, 1961; Parsons & Shils, 1951; Schwartz, 1991), have reduced the conceptualization of culture to only a few (e.g., four or five) value dimensions. These culture perspectives, focused largely on values that vary across cultures, provide an integrative system of classification that continues to be applicable to a range of both basic and applied areas of cultural research. For instance, Hofstede's culture taxonomy has provided much guidance for researchers who have sought to identify value differences among nations. As such, these taxonomic systems have provided a basis for many fundamental advances in the social sciences.

Despite the contribution of existing taxonomies of culture, researchers increasingly criticize a reliance on the simplistic values approach (Sivakumar & Nakata, 2001). In addition, scholars have also noted that existing taxonomies of culture do not capture the malleability of culture over time or within-country cultural heterogeneity (Fu, Chiu, Morris, & Young, 2007). More indirectly, other researchers suggest that cultural differences and their influence on behavior may be explained by constructs and mechanisms other than cultural values, attitudes or beliefs (Gelfand, Erez, & Aycan, 2007). To this end, the narrow focus on these values alone as an influence on cross-cultural behavior may also be considered a limitation.

2.1 Moving Beyond the Traditional Value-Based Approach: The Consideration of Context

Although culture can be shared among individuals within the same community, there is growing recognition that culture can also be fragmented and malleable within communities and across contexts as well (DiMaggio, 1997). Scholars have proposed that individuals may have several cultural perspectives or frames from which to interpret and respond to sources of influence in the environmental context. For instance, Chao and Moon (2005) use the term 'cultural mosaic' to describe the multicultural heritage of individuals that is as varied as a mosaic tile. They argue that demographic (i.e., inherited characteristics), geographic (i.e., regional characteristics), and associative affiliations (i.e., formal and informal groups with which one identifies like family, religion, or politics) shape and comprise a person's culture. Chao and Moon use the term 'cultural identity complexity' to describe the multiple cultural identities that people possess, noting that multiple and different aspects of one's identities (or tiles) can be activated at any given time depending on the sources of influence in the environmental context.

Triandis (1972) also suggests that differences in the behaviors of individuals from different cultural backgrounds are due to how subjective, observable features of culture guide behavioral intentions. Empirical studies provide further evidence that particular social and physical situations can trigger and facilitate switching between cultural lenses or frames (LaFromboise, Coleman, & Gerton, 1993). Other researchers demonstrate the conditions under which one's cultural identity can be activated through the use of external stimuli (Hong, Morris,

Chiu, & Benet-Martínez, 2000). Across these numerous studies, it is evident that one's culture, shaped by a plethora of social influences, serves as a lens or frame through which individuals perceive, interpret and respond to external cues in their environment (Hong et al., 2000; LaFromboise, Coleman, & Gerton, 1993).

Although the focus of this chapter is not on the means by which aspects of the environment guide behavior, researchers have begun to make great strides in identifying how sources of influence in the environmental context, such as socially accepted scripts and roles, can constrain and shape differences in behavior across cultures. Some sociologists suggest the cultural resources in the environmental context can be used by actors when deciding how to act, proposing that culture can be defined as a "repertoire" of techniques (Tilly, 1992) and a "toolkit" of strategies (Swidler, 1986). Just as cultural beliefs can affect behavior in the form of rituals and practices, Swidler suggests that variation in the means by which cultures organize, interact, regulate conduct, and engage in ritual practices depends on the use and availability of habits, skills and styles. She argues that when habits, skills and styles are broadly available and social norms of appropriateness permit, people have more freedom to utilize them in a variety of idiosyncratic ways to express their identity, emotions, and goals.

In sum, traditional views of culture tend to categorize culture into a set of internalized beliefs, attitudes and values. However, more recent theory has begun to conceptualize cultural values, attitudes and beliefs as more complex and malleable and has also started to incorporate the idea that culture can also be influenced by the environmental context. Given that examples exist regarding how actions and responses may be constrained or accelerated by the situation, we advocate the importance of exploring other sources of influence that exist in the external environment. In the following section, we address the taxonomic efforts that are somewhat limited in scope by discussing both the intra-individual and external factors that shape psychological processes and behavior across cultures in our framework.

3   A Contextualized Framework of Culture

Our framework is designed to address many limitations of prior taxonomies. Following the criteria of Fleishman and Quantinance (1984) for developing a viable classification system, we have selected variables that theoretically help describe actors across cultures and the environment in which they are embedded. We focus both on how values, attitudes, beliefs as well as the context comprise one's cultural identity. As previous researchers have also done, we expand the values perspective to consider how other psychological functions, such as cognition, affect and motivation can also influence behavior. However, we provide a framework that can support future research that seeks to examine how these processes occur given additional influences from the local, regional, and national context. To account for the influence of various stimuli in the contextual environment that may independently or

conjointly shape behavior, we not only outline internal values, attitudes, and beliefs in the section below, but also provide a means of broadly classifying features of the environmental context.

Through this framework, we aim to advance studies that focus on our understanding of how differences in psychological responses and behaviors across nations and cultures can be explained by considering contextual variables. Our efforts are aimed not only to assist scholars who seek to identify the mechanisms that explain or predict differences in psychological processing and behavior across cultures, but the work of scholars who also seek to simulate and describe human behavior across a variety of settings. For instance, recent research in psychology that relies upon a situated sampling method involves examining cross-cultural differences across distinct populations in the context of a representative set of social situations (Kitayama et al., 1997; Morling et al., 2002). Other research, conducted by computational modelers, aims to simulate and forecast behavior by placing actors or computer agents in a variety of virtual, man-made contexts that mirror real-world contexts. In both of these research domains, our framework provides a means to systematically develop novel hypotheses and thus drive future studies about the interaction of values, attitudes, beliefs and the aspects of the context on subjective responses and objective behaviors.

3.1 Development of a Contextualized Cultural Framework

To develop our framework, we first conducted a thorough multidisciplinary review of current and relevant cultural and social theory across several domains, including psychology, management, sociology, anthropology, and communication. Theories, taxonomies, and frameworks from each of these disciplines were reviewed to determine (1) their approach to explaining culture (e.g., taxonomy, framework, lens, internal states, external cues); (2) relevance to describing actors from different cultures and cultural contexts; (3) their level of analysis (e.g., individual, team, society); (4) validity (e.g., amount of support for this approach); and (5) ability to measure the cultural characteristics described. This information allowed us to determine the relevance of such theories to the framework as well as to begin integrating different aspects of such theories.

We then identified common themes to categorize and classify dimensions of culture, including values, attitudes, beliefs and features of the external local, regional or national context that may also explain cross-national differences. Based on our review, we classified psychological processes that may affect culture into the three categories of cognition, motivation, and affect. Specifically, we believe these three categories guide attention, information processing, and behavior, depending on how the individual or collective derives meaning from the social or environmental cues. In particular, we discuss how existing cultural dimensions, such as those by Sutton, Pierce, Burke, and Salas (2006), Schwartz (1992), and Hofstede (1980a, 1980b), are appropriately classified in one of these supra-categories of cognition, motivation,

and affect. This classification system is also broad enough to include additional dimensions of culture beyond those presented here in this framework. Together, the dimensions included in our framework can be used to differentiate among communities across the globe. The following presents a high level overview of the resulting supra-categories that address the values, affect, and motivations, along with some of the proposed factors that we believe comprise each. Furthermore, we present a classification upon which external context can be considered in conjunction with these supra-categories.

3.2 Internal Sources That Influence Culture

3.2.1 Cultural Values

People are thought to perceive, process, retrieve, and respond to information through the reliance on schemata or mental models. Default assumptions about people, objects, and events and their relationships are stored as cognitive representations. Representations of the self (Markus & Kitayama, 1994), events (Garfinkel, 1987), roles (Fiske & Taylor, 1991), and team process (Cannon-Bowers, Salas, & Converse, 1990; Mohammed & Dumville, 2001) have been empirically tested and proven to guide behavior and social interaction through their influence on expectations and biases about the appropriate ways of acting. Although some representations may be largely universal, others are culturally idiosyncratic, creating variation across groups. Cognitions about the self, interaction with others and beliefs about how to approach social situations include the dimensions of individualism-collectivism, locus of control, power distance, masculinity-femininity, and long-term vs. short term orientation. Each of these are further delineated below.

3.2.2 Individualism-Collectivism

Individualism and collectivism are probably two of the most well-known words in the study of culture. This dimension describes the relationship between the individual and the group in terms of which unit is more important to the individual in question (Sutton et al., 2006). Specifically, these values differ with regards to the relative interdependence or independence felt towards one group (Markus & Kitayama, 2004). For example, individualists tend to be more independent and focus more on themselves than the group. They promote an "I can take care of myself" attitude, are comfortable working alone, and tend to speak out and share their thoughts quite openly. Generally, they tend to distinguish themselves from others based on the unique traits they possess (Markus & Kitayama). On the other hand, collectivists put the needs of the group above their own, and value loyalty and harmony. Furthermore, collectivist individuals tend to define themselves by traits that are also characteristics of the group to which they belong (Markus & Kitayama).

Research also provides further understanding of the constructs of individualism and collectivism. For instance, Triandis and Gelfand (1998) illustrate how both individualism and collectivism may be horizontal, emphasizing equality, or vertical, emphasizing hierarchy. Most published research tends to focus on the former, which has constrained our understanding of how vertical horizontalism and collectivism may affect a variety of important social phenomenon such as leadership, justice, and conflict. To date, researchers have identified how people who are horizontally and vertically collectivist responded to the allocation of rewards differently (Chen, Meindl, & Hunt, 1997) and that vertical oriented cultures place more value on achievement and display of success than horizontally oriented cultures (Nelson & Shavit, 2002). Other researchers have also established individualism and collectivism as mediators of differences between cultures by using measures that are culture specific (Matsumoto et al., 1997).

3.2.3 Locus of Control

This dimension captures the degree to which results are perceived to be due to factors inside (internal locus of control) or outside (external locus of control) an individual (Rotter, 1966). The expectations about whom or what is responsible for events or outcomes are thought to develop from one's cultural, religious, and familial affiliations and past experiences. For example, if individuals or collectives believe fervently in God's will or other external force(s), they are likely to have a stronger external locus of control. In contrast, if collectives believe that their own skill and effort determines outcomes, they are likely to have a strong internal locus of control. Along related lines, fatalism, the belief that whatever happens must happen (Bernstein, 1992) or that one does not have full control over one's actions is also thought to vary across cultures (Aycan et al., 2000). Hence, psychological processes and behavior may be shaped by expectations about outcomes and will differ across local, regional, and national boundaries, depending on the collective's locus of control and fatalism.

3.2.4 Power Distance

Power distance is defined as the extent to which members of institutions and locations expect and accept that power is distributed equally (Hofstede, 1991). In low power distance cultures, there is believed to be minimal emphasis on the obedience of subordinates to their superiors, whereas in high power distance cultures, this is not the case (Hofstede, 1980a, 1980b, 1991). Furthermore, in a lower power distance culture, members may speak up and listen to others, despite differences in status. This, in turn, is likely to promote the willingness of subordinates to approach and question those with higher status and power (Hofstede, 1991, 2001a).

3.2.5 Masculinity–Femininity

According to Hofstede and Hofstede (2005) a society is called masculine "when emotional gender roles are clearly distinct." In contrast, a society that can be characterized as feminine is one where "emotional gender roles overlap" (Hofstede & Hofstede, p. 120). Masculine values include being assertive, having a high degree of ambition and desire for material things (Hofstede, 1980b). In contrast, femininity is defined as the opposite with values that preference caring, cooperation, and a friendly atmosphere (Hofstede, 2001a). The degree to which masculine values prevail over feminine values, determine the masculinity of a culture. Similarly, femininity is based on the extent to which a culture values feminine values more than masculine values.

3.2.6 Long-Term vs. Short Term Orientation

This dimension describes how different cultures view time (Hofstede & Bond, 1988; Sutton et al., 2006). Long-term orientation is defined by the value placed on future-oriented values such as persistence and saving. In contrast, a short-term orientation refers to values that are focused on respect for tradition and fulfilling social obligations that are oriented towards the past and present. In addition, Hofstede and McCrae (2004) suggest that focus on the short term can also translate into efforts to protect one's "face" or reputation in Asian cultures. Other research demonstrates that long-term orientation was the best predictor of country differences in norms governing emotional expressivity (Matusmoto et al., 2005) and emotional experience (Matsumoto, Nezlek, & Koopman, 2007). In the following section, we further explore how emotion can vary across cultures.

3.3 Affective Dimensions

Scholars believe that emotions enable people to adapt to environmental challenges and opportunities (Arnold, 1960; Frijda, 1987). Moreover, Hochschild (1979) proposed that emotions can and are subject to acts of management. An individual can choose to exert effort to induce or inhibit feelings so that his or her emotional expression is considered appropriate in a particular situation. To determine the appropriateness of emotional expression, culturally prescribed scripts and roles provide the rules for managing feelings (Hochschild, 1979; Rafaeli & Sutton, 1987). Hence, cultural guidelines for the expressivity and control of emotion may vary across individuals and change given the dictates of a particular situation. Furthermore, variation in the interpretation of emotional display may also be explained by distinct cultural frames of understanding due to differences in attribution.

3.3.1 Emotional Expressivity and Control

Although emotion is universal across humans, the expression, control, and perception of emotions by society is influenced by culture. Norms can exist about the valence of emotions and when they are appropriate to be expressed. Culture can also shape beliefs about how one should feel when particular situations occur and how one should act when he or she feels particular emotions (Matsumoto, 1989). The emotional expressivity dimension captures the extent to which an individual is likely to externally display affect, which can be vocal, facial, or gestural (Gross, 1998; Gross & John, 2003). The emotional control dimension captures the way we manage or control emotional impulses (Watson & Greer, 1983). Matsumoto and Ekman (1989) explored the differences in agreement among American and Japanese samples about their perception of universal emotional expressions, finding that Americans gave photos of happy and angry expressions the highest intensity ratings, while Japanese gave photos with disgusted expressions higher intensity ratings. Studies also suggest that there are cultural differences in self-reported emotional experiences, including intensity, duration and control for emotion (Scherer, Matsumoto, Wallbott, & Kudoh, 1988). Given the demonstrated relationship between emotion and culture, we have included emotional expressivity and control to our framework.

3.4 Motivational Dimensions

Differences in external and internal sources of motivation are likely to create variation in responses to environmental stimuli across cultures. Intrinsic motivation is the engagement in actions for their own sake with the only tangible benefit being outcomes such as pleasure, learning, satisfaction, interest, or challenge. Extrinsic motivation is the engagement in activities for the purpose of attaining rewards, such as praise, and to avoid punishment, such as fines (Alderman, 2004; Deci, 1976). Across cultures, differences in self-construal and what people value can predict how individuals are motivated behave. For instance, people with an interdependent self are likely to put a higher priority on the preferences of others compared to people who have an independent self-construal and this difference has been shown to affect how people prefer to make choices. Similarly, motivations to prevent negative consequences have been found to characterize individuals with interdependent selves, whereas the desire to achieve is more common among individuals with more independent self-construal (Heine et al., 1997; Lee, Aaker, & Gardner, 2000). Iyengar and Lepper (1999) also found personal choice to be more intrinsically motivating to people with independent self-construal, while people with an interdependent self-construal preferred to have decisions made for them by trusted peers. These studies shed light on the cross-cultural differences that may predict what individuals are motivated to do, choose, and be.

3.4.1 Self-Enhancement and Self-Criticism

Research suggests that people in the West have a tendency and motivation to maintain and enhance an overall evaluation of themselves or to sustain their self-esteem (Solomon, Greenberg, & Pyszczynski, 1991; Taylor & Brown, 1988; Miller & Tesser, 1989). At the same time, people in other cultures, such as Asian cultures, have been shown to be more sensitive to negative self-relevant information that can indicate whether one has fallen short of meeting the standards of excellence of their social unit, such as one's family or work group (Kitayama & Karasawa, 1997). This is evidence of how goals and motivation can vary based on the values that are characteristic of the culture in which people are embedded (Gelfand & Dyer, 2000). Along the same lines, Heine and Lehman (1997) examined whether there would be differences in the experience of cognitive dissonance after making a difficult choice among Japanese participants, who had a more interdependent self-construal, and Canadians, who tended to have a more independent self-construal. These researchers found that the possibility that one has made a bad choice does not affect individuals with an interdependent self-construal as much as it did people with independent self-construal because this action and consequence was not a threat to their sense of self, which is predominately determined by their interpersonal harmony with others, rather than their own personal achievements. Hence, self-enhancement needs were not found to be universal across these populations.

3.4.2 Uncertainty Reduction

Uncertainty avoidance is one form of uncertainty reduction that can be defined as the extent to which members of a culture feel threatened by ambiguous or unknown situations (Hofstede & Bond, 1984). In uncertainty avoidance cultures, people are likely to shun ambiguous situations (Hofstede & Hofstede, 2005). These cultures are likely to engage in actions, either risky or safe, to reduce feelings of anxiety or discomfort associated with the ambiguity. In contrast, countries or societies with weaker uncertainty avoidance are more comfortable with ambiguous situations; waiting or engaging in activities without rules or guidelines is acceptable. In countries shaped by political and social conflict, such as war, where the population is characterized by a high degree of uncertainty avoidance, anxiety levels may be higher. This results in less tolerance and willingness to wait for resolution. Thus, various factors may influence one's motivations to reduce ambiguity, but his or her level of uncertainty avoidance can influence the decision to initiate, persist, or halt a course of action.

3.5 External Sources That Influence Culture

We expand the use of our framework of cultural dimensions to also include a strategy for classifying and gathering data about the aspects of culture in the environmental context. We propose a set of dimensions that can be used independently and in combination to characterize and classify the external environment. These dimensions encompass the culturally shaped ways that geography, language and physical and social events define the context. The specific environmental dimensions used, and the rational supporting them are described below.

3.5.1 Social Context

The social context of a country consists of components such as family structure and religion. Cultural meanings about the roles of men and women, husband and wives, children and parents tend to be reproduced and reinforced through customary social interaction (Triandis 1989). Beliefs about how people should behave with others in various situations are guided by the ways that different cultures deem it appropriate to behave as a married couple, as father and son, or as friend to friend. Todd (1985) illustrated differences in family structures across countries. He used various criteria, including whether sons live independently after marriage, means by which inheritance is divided, and the acceptable marriage arrangements between cousins to develop a classification of eight family structures that differ across cultures. These classifications included the following types of family structures: patrilineal, matrilineal, monogamous and polygamous. The influence of the types of family structures across countries has been shown to also affect the way that we interact with superiors at work. For instance, Aycan et al. (1999) identify how paternalistic relationships between subordinate and superiors in the workplace characterize some countries more than others, ultimately predicting cross-cultural variation in work processes and outcomes.

Religions vary in the degree to which they are present and practiced across nations. Hence, religion is one factor that may also influence cross-cultural differences in thinking, feeling and behaving (Georgas & Berry, 1995). The impact of religion on psychological processes and behavior may not only be a result of how often individuals engage in religious practices, such as attending religious services, but in the degree to which religion shapes political procedures and processes (Matsumoto & Hee Yoo, 2006). For instance, in some countries, like the United States, religion and government are kept separate, while others are not. For this reason, religion may have a stronger influence on the culture of a country in some nations as compared to others. Specifically, religious values, such as fatalism or determinism, may influence how people engage with one another and respond to stimuli in their environment across cultures.

In some countries, the liberty to choose from a set of possible actions or responses in a social or situational context may be more limited than in others due to the rigidity of cultural norms (Gelfand et al., 2006). In contrast, other cultures may have more freedom to act in a variety of ways because cultural norms and sanctions are less prohibitive. Scholars suggest the strength of social norms within a society and the severity of sanctioning for breaking norms may predict the extent to which actors are likely to deviate from culturally normative behaviors (Gelfand et al.). Gelfand and colleagues (2011) show that the range of use of culturally-shaped habits, skills, and styles enacted in everyday routines and practices depend on the level of tightness or looseness of the social norms across cultural contexts.

3.5.2   Geography

Chao and Moon (2005) argue that geography can also influence the cultural orientation of individuals. They describe geographic dimension of the environment as aspects of the physical context such as climate, temperature, distance from the ocean, and whether the area is rural or urban. Consistent with previous research demonstrating significant correlations between measures of climate and cultural values (Hofstede, 1980a, 1980b), Chao and Moon posit that geographic and national resources help shape the development and form of civilizations. Recent empirical research has been found to support this idea. For example, Vliert, Huang, and Parker (2004) compared people in locations distinguished by their climate (hotter and colder) in terms of altruism. They found that people who are have more wealth in cold climates tend to be less altruistic than those who are poor; however, the reverse was found in warmer climates. This view is further supported by scholars who demonstrate that the agricultural or industrial context of a community can influence the cultural values of individualism and collectivism (Kluckhohn & Strodtbeck, 1961) and that the organization of political groups emerge differently depending on whether the community is a hunter, gatherer, agrarian or industrial society (Bauböck, 1998). For example, natural resources in the physical environment can influence the types of labor and occupations that characterize the local economic activities. The nature of work, along with labor conditions, can influence the structures of professions and occupational tasks (Triandis, 1972).

3.5.3   Language

The local language can also influence cross-cultural behavior. In particular, Nisbett and Norenzayan (2002) suggest that differences in language can account for differences in cognition, which can, in turn, affect behavior. Findings from their research provide evidence that language can influence the ways that people attend to objects and the relationships they perceive to be occurring between them. Linguists have clustered languages around the world into families based on common sounds, syntax, and lexicon and have identified differences in languages within and across

nations. The linguistic distance between two cultures can be discerned by locating the distance between these cultures on a map of these linguistic families and clusters.

3.5.4   Political Context

The political system of local, regional or national contexts may shape the psychological processing and behaviors of people from different nations. Political scientists have identified various characteristics that differentiate one political regime from another and the factors that may lead to the use of various political systems around the globe (Bueno de Mesquita & Siverson, 1995). Forms of government across cultures are numerous, but political scientists have identified eight broad types of government including, but not limited to full presidential republics, parliamentary republics, absolute monarchies, and single-party states. These political states vary in the extent to which participation of the population in developing policy is encouraged and allowed. Predictors of political structure and political decision-making have been linked to differences in collectivist versus individualist orientation and to orientations regarding power (Schmitter, 1984). Research also demonstrates how the procedures used within one's national government can also shape every day decision making. Specifically, Earley (1999) provides initial qualitative evidence of how American teams opt to use more democratic techniques, such as polling, whereas other countries leverage more authoritarian techniques, such as discussing the views of the leader to reach collective consensus.

3.6   Types of Cross-National, Cross-Cultural Studies

In their various review of the cross-national and cross-cultural organizational behavior research, scholars have categorized studies into two types (Kirkman et al., 2006; Lytle, Brett, Barsness, Tinsley, & Janssens, 1995; Tsui et al., 2007). These authors identify what they call 'Type 1' and 'Type 2' studies of culture. Type 1 studies are characterized by work that considers the role of culture as the independent variable, whereas Type 2 studies tend to consider culture as a moderating variable. In type 1 models, culture is largely conceptualized as internalized values, attitudes, or beliefs. Type two models were also more likely to use cultural values or nation as a proxy for culture. A few of the type II studies in the literature do measure perceptions of contextual differences such as perception of organizational political (Fu et al., 2004) or variables such as national wealth and social security (Huang & Van de Vliert, 2003).

We encourage more type II studies that consider the context variables that distinguishes local, regional and national cultures from one another, as moderators between cultural values, attitudes, and believes and outcomes, such as behavior (see Fig. 1). The consideration of contextual variables as moderators may help to
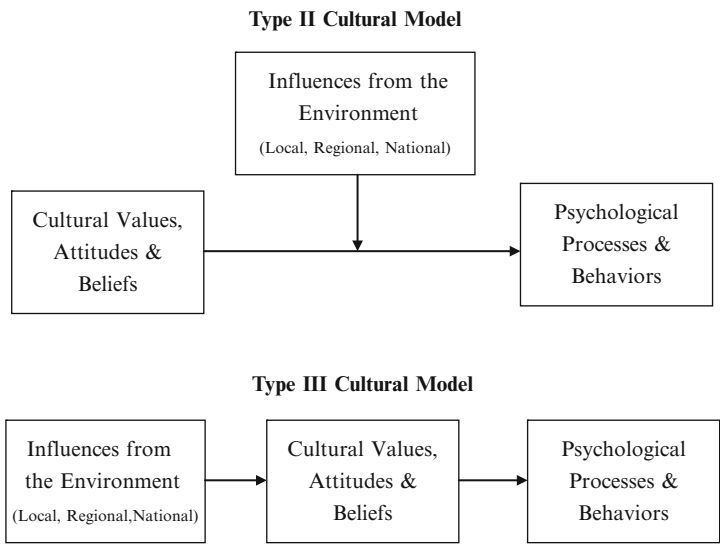
Type II Cultural Model

Influences from the
Environment
(Local, Regional, National)

Cultural Values,
Attitudes &
Beliefs

Psychological
Processes &
Behaviors

Type III Cultural Model

Influences from
the Environment
(Local, Regional,National)

Cultural Values,
Attitudes &
Beliefs

Psychological
Processes &
Behaviors

Fig. 1 Type II and III cultural models

gain a richer understanding of when cultural differences will affect psychological processing and behaviors across nations and collectives. We also propose a new type of cross-national, cross-cultural study that explores cultural values, attitudes, and beliefs as mediating variables that explain differences between collectives of individuals embedded in environmental contexts that differ with regard to their political, social, geographic or language characteristics. We illustrate this model in Fig. 1 and suggest that studies that use this type of study may help to further understand how contextual factors shape the cultural values, attitudes, and beliefs that distinguish communities across the globe.

4 Measurement Issues

Although Kitayama (2002) presents possible limitations of the dimension approach, we suggest that particular characteristics of our approach make it amenable to the construction of sociocultural simulation models that can describe and predict human, social, and cultural behavior interactions. Particularly, the breadth of existing literature using the dimension approach makes valid and reliable metrics readily available for the measurement of culture along particular dimensions of culture at various levels of analysis (e.g., individual, team, organization, or nation). Describing culture as dimensional affords this capacity as it reduces the overwhelming complexity of culture to a traceable number of factors (Hambrick, Davison, Snell, & Snow, 1998). Moreover, by focusing on cognition, affect, and

motivation of individuals and collectives and how they influence processing of external information and shape psychological processes and behavior, we hope to further enable the study of culture.

Many of the metrics associated with measuring the value dimensions included in the framework are generalizable and applicable to different levels of analysis. Hence, a criterion for the inclusion of particular values, attitudes and beliefs in our framework was whether we could measure the cultural characteristic in populations using a scale that is proven to be valid, reliable, and generalizable. Measurement of the aggregated cultural values of individuals within teams, organizations, or nations provides information about where the social unit falls along a cultural continuum on a scale. Consideration of the different contextual factors that may shape psychological processes and behaviors across nations and cultures enables empirical testing that can help us to better understand how these contextual variables shape outcomes across settings. Moreover, measurement of values, attitudes, and beliefs using a Likert scale also provides opportunities for these continuous measures to be used by computational modelers who can leverage such metrics for their computer-based games, simulations and virtual media cultural competence training studies. This will allow for development of training platforms that incorporate effective computer-based training design with sound, scientifically-driven content.

Empirical research that incorporates context, levels of analysis, and internal values into their hypotheses and design requires careful planning and consideration. At different levels of analysis, the contextual influences may have a different influence and it is important to capture this interaction. In particular, Chao (2000) advocates the idea that the conceptualization of culture at different levels can provide insight regarding cultural constructs in terms of whether such constructs are truly universal in nature or are confined to a particular collective that is exposed to certain environmental contexts or internal mental structures. For example, it may be the case that at the individual level, power distance matters more to members of a particular region due to the environmental factors impacting that region, whereas members in a different region may be more influenced by individualism. However, when aggregating individual perspectives up across these regions (such as through using an average) to obtain information regarding the national culture, such nuances may not be apparent. Therefore, it is important to understand these issues of levels of analysis, particularly in terms of different types of aggregation outside of sums or averages that may be better equipped to provide accurate representations of cultural information.

We also encourage researchers to consider appropriate techniques for characterizing the culture of social groups, such as teams and nations. Typically, culture variables are represented as compositional in nature, as they are most often represented at higher levels through the use of sums or averages (e.g., Hofstede, 2001a). However, this does not necessarily mean that composition best represents cultural variables. Indeed, it may very well be that cultural variables could be more accurately represented through compilation as the predictive power of team composition, for example, may lie in differences, rather than similarities These compilation type variables can be captured through the use of standard deviations or

variances, minimum/maximum scores, or patterns (see Kozlowski & Klein, 2000), and may allow for factors such as diversity in cultural beliefs or behaviors to be more appropriately represented. When conceptualizing culture as a multilevel issue, it is important to not only assess the most suitable level of analysis for the cultural issue of interest, but also to consider the way in which to represent culture as an emergent construct so that the most accurate representation is selected.

To measure features of the external context included in this framework, we encourage researchers to measure perceptions of features of the environmental context or to gather data that directly measures some feature of the context, such as the location or type of political regime. Ethnographic and observational methodologies from anthropology and sociology would also be most appropriate. Through field research, qualitative data can be captured that explains how patterns of interaction and the meaning attributed to routines, roles, phrases, and objects vary across cultures. Drawing from methodologies of political scientists, anthropologists, theologians, and geophysicists may also provide new techniques for characterizing the environmental contexts that distinguish communities around the globe. The consideration of the influences external to individuals on how people process information and behavior will provide experimental and computational researchers with various contextual factors in which to study differences in behavior that can complement the work of psychologists that has tended to focus on intra-person cultural values.

5 Discussion

Overall, this framework is designed to (1) provide key definitions and terminology that facilitates shared understanding of cultural dimensions; (2) describe the relationship between these dimensions and contextual features across multiple levels of analysis; and (3) facilitate the generation of hypotheses for future studies. By incorporating values, attitudes, beliefs and aspects of the external environment into our framework, we are able to conceptualize culture in a contextualized manner. In doing so, we bridge a gap between existing cultural perspectives that rely largely of the cultural values perspectives and recent advances in cultural research that argues that context matters and should be taken into account as well (Gelfand et al., 2006; LeVine, 1984; Morris et al., 2000; Triandis, 1972). Characteristics of our framework, including our variable selection criteria and the ease of measurement of values, attitudes, and beliefs also make the framework applicable to multiple levels of analysis. Finally, our framework of culture motivates the importance of considering how features of the contextual environment may suppress, activate or reverse cultural behaviors across local, national, regional communities. We provide a means to begin to systematically understand the relationship that the sources of influence in the environmental context has on cross-cultural behavior. Advanced in any of these areas can help inform computational modeling of cultural differences, aiding in training and development of cultural competence.

Certainly, the complexity of culture can be difficult to capture in a single framework. Given the multiple ways that behavior is influenced across cultures, our framework of culture takes a first step at helping to guide future studies of culture by bridging disparate views on how cross-cultural behavior is shaped. By considering the mutual influence of values, attitudes, motivation and the external context on cross-cultural behavior, we extend the work of existing cultural taxonomies and frameworks. More specifically, we contribute to the field of cross-cultural research by accounting for the cognitive, motivational, and affective dimensions of culture that shape how one perceives, processes, and responds to information, while also taking into consideration features of the environmental context that might shape the activation of these psychological processes and behavioral outcomes.

From a theoretical perspective, this taxonomic categorization of values, attitudes, beliefs and the sources of influence in the external context provides a rich set of variables that researchers can consider when generating hypotheses about behavior across cultures. A theoretical contribution must contain the essential components of *who*, *what*, *when*, *where*, and *how* (Whetten, 1989). Cross-cultural researchers have empirically and theoretically provided evidence that demonstrates how values, attitudes, and beliefs distinguish people from different local, national, and regional contexts and how these psychological states shape behavior. Future studies that consider how values, attitudes and beliefs may interact within and across individuals and communities to influence behavior could be developed from the use of the framework presented in this chapter. In addition, studies that examine the conditions under which these values shape behavior will further advance our understanding of cross-cultural differences in behavior. Since meaning is often believed to be derived from context (Gergen, 1982), attempting to understand why people behave in a particular fashion across families, teams, or countries, requires the systematic investigation of the relationship between values, attitudes, beliefs and behavior within culturally-relevant contexts. Hence, research focusing on the role that features of the external environment play in the relationship between culturally driven psychological states and behaviors has the potential to greatly advance our understanding of cross-cultural differences by providing insight about external triggers that shape behavior.

From a practical perspective, there are benefits to taking a more holistic approach to understanding culture. First, the present framework provides a systematic methodology towards inclusion of context, an area lacking in much existing cultural competency training today. By providing a guiding framework that addresses the critical variables influenced by culture, our framework can be utilized by trainers when developing scenario-based educational tools that incorporate cultural data into the design of scenarios, agent behaviors, and the virtual environment. Developers can systematically assess and design environmental characteristics, such as the social context, physical geography, and demography of the region of interest that may improve the quality of training scenarios. Doing so, may avoid the tendency to develop a culturally rich environment that does not advance training due to the limited understanding of the relationship between the environmental context, values, attitudes, and beliefs and their mutual impact on behavior. Finally, the

inclusion of variables that can be measured quantitatively may also be useful to programmers and trainers who can incorporate this data into their code when developing training and assessment tools. In sum, we believe that this framework will enable computer programmers and trainers to better conceptualize how the features of the environmental context may impact values, attitudes and beliefs (and vice versa).

6   Conclusion

We encourage the use of this cultural framework when studying psychological processes and behavior that vary across cultures because it takes into consideration both sources of influence in the external environment and cultural values, attitudes and beliefs. We hope the perspective that we have presented in this chapter stimulates continued conversation about the influence of the environmental context on psychological processes and behavior, in addition to and in combination with the influence values, affect and motivation that tend to vary across cultures. Research that adopts a contextualized approach will expand our current views of culture by identifying the conditions under which values, attitudes, and beliefs and the contextual environment shape behavior differently across cultures. Additionally, we are confident that our framework will facilitate continued dialog between social scientists and computational modelers who seek to develop sound, theoretically driven and systematic methods to quantify culture in efforts to train cultural competency and forecast cultural differences across populations.

References

Alderman, M. K. (2004). *Motivation for achievement: Possibilities for teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Arnold, M. B. (1960). *Emotion and personality* (Vols. 1 and 2). New York: Columbia University Press.

Aycan, Z., Kanungo, R. N., Mendonca, M., Yu, K., Deller, J., Stahl, G., et al. (2000). Impact of culture on human resource management practices: A 10-country comparison. *Applied Psychology: An International Review, 49*, 192–221.

Aycan, Z., Kanungo, R. N., & Sinha, J. B. (1999). Organizational culture and human resource management practices: The model of culture fit. *Journal of Cross-Cultural Psychology, 30*(4), 501–526.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bauböck, R. (1998). Sharing history and future? Time horizons of democratic membership in an age of migration. *Constellations, 4*(3), 320–345.

Bernstein, M. H. (1992). *Fatalism*. Lincoln, NE: University of Nebraska Press.

Busenitz, L. W., West III, G. P., Shepherd, D., Nelson, T., Chandler, G. N., & Zacharakis, A. (2003). Entrepreneurship research in emergence: Past trends and future directions. *Journal of Management, 29*(3), 285–308.

Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Cannon-Bowers, J., Salas, E., & Converse, S. (1990). Cognitive psychology and team training: Training shared mental models and complex systems. *Human Factors Society Bulletin, 33*(12), 1–4.

Chao, G. T. (2000). Multilevel issues and culture: An integrative view. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 308–348). San Francisco: Jossey-Bass.

Chao, G. T., & Moon, H. (2005). The cultural mosaic: A metatheory for understanding the complexity of culture. *Journal of Applied Psychology, 90*(6), 1128–1140.

Chen, C. C., Meindl, J. R., & Hunt, R. G. (1997). Testing the effects of vertical and horizontal collectivism: A study of reward allocation preferences in china. *Journal of Cross-Cultural Psychology, 28*(1), 44–70.

Deci, E. L. (1976). Notes on the theory and metatheory of intrinsic motivation. *Organizational Behavior and Human Performance, 15*(1), 130–145.

De Mesquita, B. B., & Siverson, R. M. (1995). War and the survival of political leaders: A comparative study of regime types and political accountability. *American Political Science Review, 4(1)*, 841–855.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology, 23*, 263–268.

Earley, P. C. (1997). Doing an about-face: Social motivation and cross-cultural currents. In Earley, P. C. (Eds.), *New perspectives on international industrial/organizational psychology* (pp. 243–255). San Francisco, CA: The New Lexington Press/Jossey-Bass Publishers.

Earley, P. C. (1999). Playing follow the leader: Status-determining traits in relation to collective efficacy across cultures. *Organizational Behavior and Human Decision Processes, 80*(3), 192–212.

Earley, P. C., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal, 43*(1), 26–49.

Erez, M., & Earley, P. C. (1987). Comparative analysis of goal-setting strategies across cultures. *Journal of Applied Psychology, 72*(4), 658.

Erez, M., & Earley, P. C. (1993). *Culture, self-identity, and work*. USA: Oxford University Press

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic.

Frijda, N. H. (1987). Emotion, cognitive structure and action tendency. *Cognition and Emotion, 1*, 115–144.

Fu, J., Chiu, C., Morris, M. W., & Young, M. (2007). Spontaneous inferences from cultural cues: Varying responses of cultural insiders and outsiders. *Journal of Cross-Cultural Psychology, 38*, 58–75.

Fu, P. P., Kennedy, J., Tata, J., Yukl, G., Bond, M. H., Peng, T. K., et al. (2004). The impact of societal cultural values and individual social beliefs on the perceived effectiveness of managerial influence strategies: A meso approach. *Journal of International Business Studies, 35*(4), 284–305.

Gabrenya, W. K., Jr. (1999). Language use in IACCP part 1: Competencies and controversies. *Cross-cultural Psychology Bulletin, 33*(3–4), 31–33.

Garfinkel, H. (1987). Studies of the routine grounds of everyday activities. In *Studies in ethnomethodology* (pp. 35–75). Oxford, UK: Polity.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Gelfand, M., & Dyer, N. (2000). A cultural perspective on negotiation: Progress, pitfalls, and prospects. *Applied Psychology, 49*(1), 62–99.

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58*, 479–514.

Gelfand, M. J., Nishii, L. H., & Raver, J. L. (2006). On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology, 91*(6), 1225–1244.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*, 1100–1104.

Gergen, K. (1982). *Toward transformation in social knowledge*. New York: Springer.

Georgas, J., & Berry, J. W. (1995). An ecocultural taxonomy for cross-cultural psychology. *Cross-Cultural Research, 29*(2), 121–157.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.

Gross, J. J., & John, O. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362.

Hambrick, D. C., Davison, S. C., Snell, S. A., & Snow, C. C. (1998). When groups consist of multiple nation-alities: Towards a new understanding of the implications. *Organization Studies, 19*, 181–205.

Heine, S. J., & Dehman, D. (1997). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin, 23*, 389–400.

Heine, S. J., & Lehman, D. R. (1997). The cultural construction of self-enhancement: An examination of group-serving biases. *Journal of Personality and Social Psychology, 72*(6), 1268.

Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *American Journal of Sociology, 85*, 551–575.

Hofstede, G. (1980a). *Culture's consequences: International differences in work-related values*. Thousand Oaks, CA: Sage.

Hofstede, G. (1980b). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics, 9*(1), 42–63.

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of International Business Studies, 14*, 75–89.

Hofstede, G. (1984). The cultural relativity of the quality of life concept. *Academy of Management Review, 9*, 389–398.

Hofstede, G. (1991). *Culture and Organizations Software of the mind*. London: McGraw-Hill.

Hofstede, G. (2001a). *Culture's consequences: Comparing, values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Hofstede, G., & Bond, M. H. (1988). The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics, 16*(4), 4–21.

Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions an independent validation using rokeach's value survey. *Journal of Cross-Cultural Psychology, 15*(4), 417–433.

Hofstede, G., & Hofstede, J. (2005). *Cultures and organizations: Software of the mind (Revised and expanded 2nd edn)*. New York: McGraw-Hill.

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*(1), 52–88.

Hong, Y. Y., Morris, M., Chiu, C. Y., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist, 55*, 709–720.

Huang, X., & Van de Vliert, E. (2003). Where intrinsic job satisfaction fails to work: National moderators of intrinsic motivation. *Journal of Organizational Behavior, 24*(2), 159–179.

Ip, G., & Bond, M. H. (1995). Culture, values, and the spontaneous self-concept. *Asian Journal of Psychology, 1*, 30–36.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and* Social Psychology, 76(3), 349.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of culture's consequences: A review of the empirical research incorporating Hofstede's cultural value framework. *Journal of International Business Studies, 37*, 285–320.

Kitayama, S. (2002). Culture and basic psychological theory—Toward a system view of culture: Comment on Oyserman et al. (2002). *Psychological Bulletin, 128*, 89–96

Kitayama, S., & Karasawa, M. (1995). Self: A cultural psychological perspective. *Japanese Journal of Experimental Social Psychology, 35,* 133–163

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin, 23*, 736–742.

Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology, 72*, 1245–1267.

Klein, H. A. (2004). Cognition in natural settings: The cultural lens model. In M. Kaplan (Ed.), *Cultural ergonomics: Advances in human performance and cognitive engineering research, 4* (pp. 249–280). Boston: Elsevier JAI.

Kluckhohn, F., & Strodtbeck, F. (1961). *Variations in value orientations*. Westport, CT: Greenwood Press.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In Klein, K. J. & Kozlowski, S. W. J. (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

LaFromboise, T., Coleman, H. L. K., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin, 114*, 395–412.

LaFromboise, T., Coleman, H. L., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin, 114*(3), 395.

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology, 78*(6), 1122.

Leung, K., Bond, M. H., & Schwartz, S. H. (1995). How to explain cross-cultural differences: Values, valences, and expectations? *Asian Journal of Psychology, 1*, 70–75.

LeVine, R. (1984). Properties of culture: An ethnographic view. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion*. New York: Cambridge University Press.

Lytle, A. L., Brett, J. M., Barsness, Z. J., Tinsley, C. H., & Janssens, M. (1995). A paradigm for confirmatory crosscultural research in organizational behavior. *Research in Organizational Behavior, 17*, 167–214.

Markus, H. R., & Kitayama, S. (1994). A collective fear of the collective: Implications for selves and theories of selves. *Personality and Social Psychology Bulletin, 20*, 568–579.

Markus, H. R., & Kitayama, S. (2004). Models of agency: Sociocultural diversity in the construction of action. In V. Murphy-Berman & J. Berman (Eds.), *The 49th Annual Nebraska symposium on motivation: Cross-cultural differences in perspectives on self* (pp. 1–57). Lincoln, NE: University of Nebraska Press.

Matsumoto, D. (1989). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion, 17*(2), 107–123.

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion, 13*(2), 143–157.

Matsumoto, D., Nezlek, J. B., & Koopmann, B. (2007). Evidence for universality in phenomeno-logical emotion response system coherence. *Emotion; Emotion, 7*(1), 57.

Matsumoto, D., Weissman, M. D., Preston, K., Brown, B. R., & Kupperbusch, C. (1997). Context-specific measurement of individualism-collectivism on the individual level the individualism-collectivism interpersonal assessment inventory. *Journal of Cross-Cultural Psychology, 28*(6), 743–767.

Matsumoto, D., & Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspectives on Psychological Science, 1*(3), 234–250.

McClelland, D. C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand.

Millar, M. G., & Tesser, A. (1989). The effects of affective-cognitive consistency and thought on the attitude-behavior relation. *Journal of Experimental Social Psychology, 25*(2), 189–202.

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89–106.

Morling, B., Kitayama, S., & Miyamoto, Y. (2002). Cultural practices emphasize influence in the United States and adjustment in Japan. *Personality and Social Psychology Bulletin, 28*, 311–323.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*, 949–971.

Morris, M. W., Podolny, J. M., & Ariel, S. (2000). Missing relations: Incorporating relational constructs into models of culture. In P. C. Earley & H. Singh (Eds.), *Innovations in international and cross-cultural management* (pp. 52–90). Thousand Oaks, CA: Sage.

Nelson, M. R., Shavit, S. (2002). Horizontal and vertical individualism and achievement values: A multimethod examination of Denmark. *Journal of Cross-Cultural Psychology, 33*(5), 439–458.

Nisbett, R. E., & Norenzayan, A. (2002). Culture and cognition. In H. Pashler & D. L. Medin (Eds.), *Stevens' handbook of experimental psychology: Vol. 2: Cognition* (3rd ed., pp. 561–597). New York: Wiley.

Parsons, T., & Shils, E. E. (1951). *Toward a general theory of action*. New York: Harper & Row Publishers.

Rafaeli, A., & Sutton, R. I. (1987). The expression of emotion as part of the work role. *Academy of Management Review, 12*, 23–37.

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1.

Scherer, K., Matsumoto, D., Wallbott, H., & Kudoh, T. (1988). Emotion antecedents and reactions in the U.S. and Japan. *Journal of Cross-Cultural Psychology, 19*, 267–286.

Schmitter, P. C. (1984). *Neo-corporatism and the state*. Florence: European University Institute.

Schwartz, S. H. (1991). Universals in the content and structure of values: Theoretical advances and empirical tests in two countries. In L. Berkowitz (Ed.), *Advances in experimental and social psychology*. San Diego, CA: Academic.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*(1), 1–65.

Schweder, R. A., & LeVine, R. A. (1984). *Culture theory: Essays on mind, self and emotion*. New York: Cambridge University Press.

Shore, B. (1996). *Culture in mind: Cognition, culture, and the problem of meaning*. New York: Oxford University Press.

Sivakumar, K., & Nakata, C. (2001). The stampede toward Hofstede's framework: Avoiding the sample design pit in cross-cultural research. *Journal of International Business Studies, 32*, 555–574.

Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior: The psychological foundation of self-esteem and cultural worldviews. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 93–159). San Diego, CA: Academic.

Sutton, J. L., Pierce, L., Burke, C. S., & Salas, E. (2006). Cultural adaptability. In C. S. Burke, L. Pierce, & E. Salas (Eds.), *Advances in human performance and cognitive engineering research* (pp. 143–173). Oxford, UK: Elsevier Science.

Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review, 51*, 273–86.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 108*, 193–210.

Tilly, C. (1992). *How to detect, describe, and explain repertoires of contention* (Working Paper No. 150). New York: Center for Studies of Social Change. New School for Social Research.

Todd, E. (1985). *The explanation of ideology, family structures and social systems*. Oxford: Basil Blackwell.

Triandis, H. C. (1972). *The analysis of subjective culture*. New York: Wiley.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review, 96*(3), 506.

Tsui, A. S., Nifadkar, S. S., & Ou, A. Y. (2007). Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management, 33*, 426–478.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74*(1), 118.

Van de Vliert, E., Huang, X., & Parker, P. M. (2004). Do colder and hotter climates make richer societies more, but poorer societies less, happy and altruistic? *Journal of Environmental Psychology, 24*(1), 17–30.

Watson, M., & Greer, S. (1983). Development of a questionnaire measure of emotional control. *Journal of Psychosomatic Research, 27,* 299–305.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University.

Whetten, D. (1989). What constitutes a theoretical contribution? *The Academy of Management Review, 14*, 490–495.

Toward a Unified Negotiation Framework: Leveraging Strengths in Behavioral and Computational Communities

Nazli Turan, Tinglong Dai, Katia Sycara, and Laurie Weingart

Abstract While there has been a large body of negotiation literature in both Behavioral Science (behavioral) and Artificial Intelligence/Game Theory (computational) communities, there has not been an attempt to bridge the two communities to our best knowledge. In this chapter, we compare and contrast the characteristics of behavioral and computational literature in negotiation. We propose that incorporating the strengths of two types of literature are valuable in expanding the horizon of research outlook.

Keywords Behavioral research • Computational approaches • Automated negotiation • Utilities • Subjective characteristics • Emotion • Agent reasoning • Negotiation literature • Interdisciplinary study • Game theory

1 Introduction

In the last few decades, negotiation has enjoyed generous research interest across disciplines from scholars who have employed methodologies and research agendas from their respective fields to unravel fundamental questions underlying bargaining situations. Political scientists have tried to reconstruct and process-analyze international negotiations between states based on official press releases from the countries (Druckman, 1986); economists have sought to understand the conditions under which negotiators equalize gains given ordinal utility scales (Myerson, 1977);

N. Turan • T. Dai • L. Weingart
Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: nturan@andrew.cmu.edu; dai@cmu.edu; weingart@cmu.edu

K. Sycara (✉)
Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: katia@cs.cmu.edu

policy-making researchers examined the role of equity for cooperation in international environmental negotiations (Lange & Vogt, 2003) computer scientists are also working on negotiation research, trying to design more intelligent automated negotiation agents (Kraus, 1997; Sycara, 1990). This consistent interest in negotiation speaks not only to the relevance of the concept in different facets of everyday human experience but also to its applicability as a field of scholarly inquiry. It is, however, intriguing that all these different research efforts in the area of negotiation have mostly remained distinct rather than capitalizing on the informative potential that lies in their convergence.

The main aim of this paper is to bridge the work of two fields that have contributed greatly to research on negotiation: behavioral sciences and computational modeling. Our aim is to delineate both the similarities and differences in how these two fields have conducted negotiation research, their main findings and future directions. To our knowledge, there has not been any prior work to date aiming at an integrated framework of behavioral and computational negotiation literature. Another important purpose of this paper, therefore, is to identify future directions of negotiation research bridging the major strengths of computation and behavior researchers.

We present our discussion and integration of the literature in terms of four aspects of negotiation research that have been examined by both behavioral and computational researchers and thus provide complementary perspectives: (1) utilities, consisting of values, motives and goals; (2) negotiator-internal states, consisting of the more affective components of the internal reasoning of negotiators such as trust, beliefs and emotions; (3) negotiator-internal reasoning, which is a cognitive process of encoding, searching and algorithm creation and (4) externally-observable behavior and characteristics such as strategies, tactics and outcomes of negotiation. However, before dwelling into these four categories, it is useful to briefly present how the two fields understand and explore negotiation in general terms. In the behavioral sciences, specifically organizational behavior, social psychology, sociology and behavioral economics, one commonly used definition of negotiation is "a form of conflict behavior, which occurs when two or more parties try to resolve a divergence of interest by means of conversation" (Pruitt & Kim, 2004, p. 56). This general and broad definition perhaps demonstrates the fact that negotiation is one of the most common yet at the same time most complex human activities (Lewicki et al., 1997). People constantly engage in negotiations in their social and professional lives to solve issues that contain both shared and opposed interests (Fisher & Ury, 1981). This means that each negotiation situation contains potential for both competition and cooperation and almost inevitably mutual interdependence (Lewicki et al., 1997). Behavioral negotiation researchers, therefore, focus on understanding the process and outcome of negotiation under various settings.

In the computational literature, there exists similar definitions of negotiation as in the behavior literature. Braun et al. (2006) define negotiation as "a decentralized decision-making process used to search for and arrive at an agreement that satisfies the requirements of two or more parties in the presence of limited

common knowledge and conflicting preferences." The research focus of most of the computational literature, however, is different in that it focuses on providing negotiation support systems (or e-Negotiation systems) to enable automated negotiations between intelligent, autonomous agents, or to design automated negotiation agents to negotiate with human counterparts, or to help and advise negotiators during the various phases during the negotiation process.

It should be noted that there is comparatively more knowledge on negotiation in the behavioral sciences since negotiation has been studied in these fields for a much longer time than it has been in computational sciences. It is also true that since research in certain behavioral sciences such as psychology or organizational behavior often uses experimental methods with human subjects, complex interactions between people can be captured. As a result, knowledge about negotiation discovered by behavioral sciences has been more complex than in the computational sciences because of the possibility to include or control for multiple factors such as individual differences, evolving structure of the negotiation, possibility of future interaction and so on. Therefore, computational sciences could take advantage of the findings from behavioral sciences to inform their models of negotiation.

2   Negotiation Research: Different Perspectives

Negotiation research has been classified using a variety of typologies. Raiffa (1982) proposes that negotiation papers can be classified into four categories: (1) symmetrically descriptive, (2) symmetrically prescriptive, (3) asymmetrically prescriptive/descriptive, and (4) externally prescriptive or descriptive. The first category of research focuses on "describing the behavior of all the negotiators, without having any interest whatsoever in prescribing how they should behave". The second category of papers provide advice regarding how negotiators should behave if they possess infinite rationality and reasoning capacities. Such advice given to all parties must be in equilibrium and no one thus has the incentive to deviate from it. The third is concerned with "studying and understanding the behavior of real people in real conflict situations, so that he can better advise one party about how it should behave in order to achieve its best expected outcome". The last category differs from all the other three in that it addresses the problem of an intervener (e.g., mediator, arbitrator, and rule manipulator) during a negotiation scenario. The research helps the intervener manage the negotiating process "in some impartial, balanced way".

Another way of classifying negotiation research is to look at whether a paper focuses on negotiation process or negotiation outcome. In the negotiation literature, a process refers to the events and interactions that occur between parties before the outcome. A process includes all verbal and non-verbal exchanges among parties, the enactment of bargaining strategies and the external and situational events that influence the negotiation (Thompson, 1997). Process analysis in bargaining has mainly focused on either the back and forth exchanges between the negotiators (e.g., Adair & Brett, 2005) or on the broader phases of strategic

activity over time (e.g., Olekalns, Brett, & Weingart, 2003). A more recent trend has been to examine the "interplay between moment-to-moment actions and reactions exhibited by negotiators within their broader behavioral/strategic context" (Olekalns & Weingart, 2008).

A negotiation outcome, on the other hand, is the "product or endpoint of bargaining" such as an agreement, impasse or deadlock (Thompson, 1997). The most general categorization that comes from such analysis of negotiation outcomes and processes is the distinction between *competitive* and *cooperative* situations, which is also referred to as *distributive* vs. *integrative* bargaining. While these terms are often used interchangeably, it is useful to note that cooperation vs. competition is used more often for negotiation processes whereas the latter is used more often to denote the structural elements of the negotiation as well as the outcomes that result from them. Competitive negotiation occurs when "the goals of one party are in fundamental and direct conflict with the goals of the other party" and where "resources are fixed and limited and thus each party wants to maximize his own profit" (Lewicki et al., 1997). On the other hand, cooperative negotiation entails that "goals of the parties are not mutually exclusive" (Lewicki et al., 1997). Thus, in cooperative processes, parties can engage in "positive moments to increase the potential gains relative to no agreement" whereas in competitive ones, they utilize moves that "threaten to increase the losses for the other party" (Walcott & Hopmann, 1974). Similarly, in distributive situation, a fixed resource is simply divided whereas in integrative situations, interests of both parties are satisfied although there may be concessions on both sides (Lewicki et al., 1997). In other words, the terms integrative and distributive are used both to describe issues, i.e. an integrative issue is one of differential importance to each party, making trade-offs possible whereas a distributive one is where interests are diametrically opposed. Similarly, an integrative outcome is reached through trade-offs on integrative issues, thus leaving both sides satisfied whereas distributive outcomes are those where a fixed resource is simply divided among parties. Thus, it is safe to say that in most cases a cooperative process leads to integrative outcomes whereas a competitive process leads to distributive outcomes. However, it is also true that it is limiting and challenging to try to divide a complex human interaction like negotiation into these two clear-cut and opposing categories. Thus, a third category, mixed-motive, is used to refer to bargaining situations where parties pursue interests which usually are competing and compatible at the same time (Fairfield & Allred, 2007).

3 The Present Framework

As suggested above, we consider the behavioral and computational research on negotiation by organizing the literature into four categories: subjective utilities, negotiator- or agent-internal states, negotiator- or agent-internal reasoning and externally-observable behavior.

3.1   Utilities

In behavioral research, subjective utilities, mainly values, goals and motives in negotiation, are constructed by the individual but are also influenced by the social context that places constraints on these preferences. One of the early models to reflect this phenomenon is the Dual Concern Theory (Pruitt & Rubin, 1986). Although originally developed as a conflict management framework, issues addressed by this model are very much in tandem with those of negotiation and are thus used very often in this arena. The model outlines five basic conflict management strategies: competition, collaboration, compromise, avoidance and accommodation. An individual's choice of the strategy to employ in a conflict situation will be determined as a product of concern for self and concern for the other, which can be called self-concern and other-concern. For example, people with high self-concern and low other-concern will tend to compete just as those high in both will collaborate or those low in self but high in other concern will tend to accommodate.

Most of the application of the self vs. other concern dichotomy in behavioral negotiation research has been within the framework of social motives or social value orientations (Deutsch, 1949; Van Lange, 1999). Although these two terms are used almost interchangeably in the literature, the latter is used more commonly to refer to differences in social motives rooted in personality traits whereas the former is used both for situational and individual differences. The main dichotomy in this framework is the prosocial vs. egoistic social motive with the egoistic social motive usually being broken down into individualistic and competitive components. When applied to negotiation, it is expected that prosocials will have the aim to maximize outcomes for both self and others whereas individualistic negotiators will seek to maximize only own outcomes and competitive negotiators will try to maximize relative advantage over others (Van Lange). In other words, prosocials draw positive value from others' outcomes whereas individualists draw none and competitors draw negative value from others' outcomes in negotiation. The implications of these propositions, especially for integrative negotiation, have been empirically tested through multiple studies. In a meta-analysis of 28 studies examining the role of prosocial vs. egoistic motives and resistance to yielding in negotiation, De Dreu, Weingart, and Kwon (2000) found that "negotiators were less contentious, engaged in more problem-solving and achieved higher joint outcomes when they had a prosocial rather than egoistic motive" when resistance to yielding was not low. Findings of the authors extend previous work by demonstrating that prosocial negotiators engage in more problem-solving behavior and resort less to contentious tactics, which is critical because it sheds light on some of the important behavioral mechanisms leading to integrative agreements.

What people value in negotiation is another broad theme that falls under the subjective utilities category. According to the *economic models of bargaining* that dominated the field in its nascent stages, the ultimate aim in negotiation is maximizing one's own outcome and the easiest and most efficient way to realize this aim is through integrative potential (Nash, 1953). However, it is now well-documented

in the field that pure economic outcomes are poor indicators of not only what people value in negotiation but also of their behavioral manifestations. Research has shown that perceptions of self, relationship with the other party or the desire to maintain a positive image may be as influential as, if not more, than economic gains. Issues such as self-efficacy, self-esteem, maintaining face or maintaining social relationships with the other party may be of critical concern to the negotiators and subsequently influence processes and outcomes (Anderson & Shirako, 2008; Bandura, 1977; McGinn & Keros, 2002; Synder & Higgins, 1988). The question of what negotiators value and how it influences their perceptions of the outcome has become a fertile area of bargaining research to the extent that Curhan, Elfenbein, and Xu (2006) developed and validated a framework to measure subjective value in negotiation. The 16-item Subjective Value Inventory (SVI) includes questions about the perception of the negotiator towards the incremental outcome in terms of satisfaction, loss/win, legitimacy; and feelings about the self, the process and the relationship. The authors also find that the SVI is a more accurate predictor of future negotiation decisions than economic outcomes, which demonstrates again that what people value in negotiation cannot be fully or accurately predicted by sole profit maximization models.

The last main sub-category to be discussed within subjective utilities is goals. At this point, it is important to recognize a possible profusion of terminology in the behavioral science perspective on negotiation. On one hand, goals are used almost interchangeably as motives and are treated in the same way that motives have been described in this paper. For example, there is considerable reference to prosocial or competitive goals in negotiation (De Dreu, 1997). On the other hand, goals are also used to refer to target or aspiration points. Most of the work that approaches goals in negotiation from a goal-setting framework would fall under this latter group. A meta-analysis by Zetik and Stuhlmacher (2002) examined the influence of goal setting in negotiation and found a strong effect for difficult goals on higher individual outcomes.

In the computational literature, few papers discuss subjective utility. Rather most of the computational literature about negotiation provides a clear definition of the utility function. Fogelman-Soulie, Munier, & Shakun (1983) developed an MDP (Markov Decision Processes) model for the problem of bilateral two-issue negotiation. Instead of assuming *bivariate utilities*, the *one-stage* payoff is expressed as a payoff probability distribution representing the probability that a player obtains various amounts of each of the two variables. Kraus, Wilkenfeld, and Zlotkin (1995) discussed different forms of continuous utility functions over all possible outcomes, e.g., time constant discount rates and constant cost of delay. Zlotkin and Rosenschein (1996) presented an approach to the negotiation problem in non-cooperative domains wherein agents' preferences over different intermediate outcomes are captured by "worth functions" by considering the probabilistic distance between intermediate outcomes and final outcomes. Rangaswamy and Shell (1997) designed a computer-aided negotiation support system, one part of which is to help negotiating parties disaggregate their own preferences and priorities in order to have a better understanding of them, utilizing several utility assessment

techniques. Faratin, Sierra, and Jennings (2002) used a given *linearly additive multi-attribute* utility function to represent agent preferences. Each agent is assumed to have a scoring function that gives the score it assigns to a value of each decision variable in the range of its acceptable values. Then the agent assigns a weight to each decision variable to represent its relative importance. Lin, Kraus, Wilkenfeld, and Barry (2008) assumed that agents have bounded rationality and their choice preferences were modeled using a utility function generated from the order of ranking of different offers. The analysis is shifted from the model of expected utility maximization to the evaluation of offers using the maximin method and the ranking of offers. To the end, the utility function definitions in the computational literature enrich the ways of modeling agent preferences under different negotiation settings.

A number of papers, however, represent the trade-off between multiple issues using constraints, i.e., the negotiators possible set of proposals, instead of utility functions. Balakrishnan and Eliashberg (1995) propose a single-issue negotiation process model where the utilities are simply the negotiation outcome, and agents' dynamic preferences are represented using a constraint with the left-hand side denoting agents' "resistance forces", and right-hand side "concession forces". Luo, Jennings, Shadbolt, Leung, and Lee (2003) consider fairness using a fuzzy constraint based model for bilateral, multi-issue negotiations in trading environments. The prioritized fuzzy constraints are used to represent trade-offs between the different possible values of the negotiation issues and to indicate how concessions should be made when they are necessary.

3.2   Agent-Internal States

Agent-internal states, consisting of the more affective aspects of negotiator behavior, such as trust, moods or emotions, have become a popular venue of research in the recent years after a long-lasting reluctance to grant them a place at the bargaining table. This growing body of research has convincingly drawn attention to the importance of these concepts in shaping not only negotiators' own but also perceptions of others' states and consequently influencing bargaining behavior (Barry & Oliver, 1996; Kumar, 1997). The field has also distinguished between different types of affective states that may be in play during negotiation (or before and after as well). *Affect* refers to the "whole range of preferences, evaluations, moods and emotions" (Watson & Tellegen, 1985, as quoted in Thompson, 1998). *Moods* are "low-intensity, diffuse feeling states that do not have a clear antecedent (Forgas, 1992)" which influence "a whole range of social cognitions and behaviors considered to be primarily positive or negative" (Thompson). Finally, *emotion* is a feeling state that has a clear cause or object, is short-lived but intense and focused (Frijda, 1994; Thompson, 1998). As can be understood from these definitions, affective states such as emotions and moods, though they certainly influence cognitions, are still understood as separate processes from them, which is a distinction that was not clearly delineated for many years in negotiation literature.

Most of the affect work in negotiation has been on the positive side and the general finding has been that positive emotions lead to a range of positive outcomes. Kramer and colleagues (1993) have found that when happy negotiators bargain with other happy negotiators, their individual outcomes are better. Carnevale and Isen (1986) replicate this finding at the joint outcome level and also find that happy negotiators perform better mainly because they engage in less contentious tactics and pressure their counterparts less towards concessions. On the other hand, there has been less research on the role of negative or neutral emotions in negotiation, owing partly to the relative difficulty of experimentally manipulating negative emotions such as anger, sadness, guilt or shame. Methods traditionally used to induce positive feelings such as making the participants watch happy videos before the negotiation do not work as effectively when applied to the negative realm. That being said, there is also considerable consensus in the field that negotiation with positive emotions lead to better outcomes than those with negative emotions (Forgas, 1998; Van Kleef, De Dreu, & Manstead, 2003). For example, angry negotiators have been found to claim more value (Anderson & Neale, 2005); to force their opponents into more concession making (Sinaceur & Tiedens, 2006) and to have outcome preferences with deleterious effects on the overall negotiation process (Loewenstein, Thompson, & Bazerman, 1989). Another point about work on negative emotion in negotiation is that most of it has been in comparative terms to negative emotions such as happy vs. sad or neutral negotiators. Therefore, the field is certainly open to expansion of knowledge how, under what conditions and through which mediators negative and neutral emotions influence bargaining situations.

In the computational literature, there is a lack of modeling efforts in terms of subjective agent characteristics such as beliefs, prejudices, emotions and cultural factors. Most of the existing models, however, clearly specify information disclosure within negotiating agents, i.e., who knows what, who understands what. Kraus et al. (1995) assume that each agent knows all relevant information about the other agent, particularly internal states and utility functions over different outcomes. Lin et al. (2008) assume that there is a finite set of agent types. Each agent is aware of the set of possible types of the opponent but not the exact utility function. The agent has some probabilistic belief about the type of the other agent and such beliefs can be updated according to the Bayesian updating rule throughout the negotiation process. Their experimental results verified that the agent in most cases believed that its opponent is of the correct type with higher probability than the opponents other possible types. Busch and Horstmann (2002) study a two-issue bargaining model with asymmetric information to study agent choice of how to structure bargaining. The problem involves a buyer and a seller who bargain over the price of two distinct goods (X and Y). The seller's valuation for each good is public knowledge; the buyer's valuation of X is common knowledge, but his valuation of Y is private information and can be either high or low. The seller updates his belief each time after he receives an offer from the buyer. The interesting part in (Busch & Horstmann) is that when the negotiation agenda becomes exogenous, i.e., when the buyer can choose whether to negotiate issue-by-issue or jointly, it is possible for the buyer to signal his type (high or low valuation of the good Y).

3.3 Agent-Internal Reasoning

The third category, agent-internal reasoning encompasses the cognitions and the subsequent behaviors that negotiators engage as they search for information with the aim of sense-making during bargaining. Commonly referred to as the cognitive approach, this perspective posits that cognitive activity can be broken down and analyzed in terms of how negotiators process and recode abstract information. Some of the most popular subjects of bargaining research such as biases, heuristics or interpersonal attributions and judgments fall under the umbrella of this perspective. Before the evolution of the cognitive approach, the field had been more interested in untangling the basic behaviors, goals and strategies of negotiators without dwelling much into their underlying processes. Building on prospect theory and behavioral decision theory, the cognitive approach in negotiation has demonstrated strong effects of biases such as availability, anchoring and overconfidence. For instance, Galinsky and colleagues (2002) find that initial offers negotiators make serve as anchors that eventually become more influential assessments of outcome satisfaction than objective results are. In a similar way, building on information-processing theory, there have been findings on the implications of how the negotiation task or process is perceived by the negotiators. All these findings show that how negotiators perceive the bargaining situation and the actors within the bargaining situation has the potential to exert process and outcome changing influence.

There exists a comprehensive body of computational literature of agent-internal reasoning. Zeng and Sycara (1998) develop an automated negotiation model wherein agents are capable of reasoning based on experience and improve their negotiation strategies incrementally. They utilize the Bayesian framework to update an agent's belief about its opponents. Lin et al. (2008) model an agent's internal reasoning in terms of generating and accepting offers. When generating offers, an agent selects the best offer among the offers that the agent believes might be accepted. To be more specific, the agent selects the minimum value of (1) the agent's own estimation of the offer and (2) the agent's estimation of its opponents' acceptable offer, under the pessimistic assumption that the probability that an offer is accepted is based on the agent that favors the offer the least. In discussing the agent's reasoning about accepting offers, they make the assumption that each offer is evaluated based on their relative values compared to the reservation price. Lue et al. (2003) express agents' preferences as "Prioritized Fuzzy Constraint Satisfaction Problems", wherein agents optimize their objective values given constraints with different priority levels. This model is the basis for the buyer agent and the seller agent to reason about generating and accepting offers during the negotiation process. Faratin et al. (2002) model agents' trade-offs between various offers/outcomes using *similarity-based trade-off reasoning*. A linear trade-off algorithm is proposed to represent agents' preferences.

3.4   Externally-Observable Behavior and Characteristics

The final category of the framework is the externally-observable behavior and characteristics such as strategies, tactics and outcomes of negotiation. Earlier we discussed how behavioral scientists conceptualize negotiation outcomes, in this section we delineate how negotiators employ strategies and tactics to reach these outcomes. For clarity of presentation, it is necessary to point out that behavioral research treats tactics and strategies differently as tactics are negotiation behaviors "employed in service of a goal" which is a very broad concept but one which is used "not in isolation but rather in combination, either consciously or unconsciously to form a strategy" (Weingart & Olekalns, 2003). With this definition, it is possible to think of how almost all the concepts that have so far been described in this framework can be utilized for tactical purposes or become part of an overall negotiation strategy. As such, negotiation researchers have examined how emotions can be used as a tactic or how perception of the other side's utility is a determinant of individual bargaining strategies. Another line of research within this category has been concerned with examining tactics and strategies within the negotiation process and how they consequently relate to the outcome as opposed to linking them directly to outcomes. An example would be process research that looks at how negotiators' tactical knowledge influences the frequency of the tactical use of this knowledge (Weingart, Hyder, & Prietula, 1996).

Computational researchers, on the other hand, have focused on agents' optimal actions based on their reasoning strategies, and the efficiency compared to Pareto optimal solutions or human negotiation outcomes. Faratin et al. (2002) present a heuristic strategy that helps agents trade off strategy between multiple negotiation decision variables (e.g., whether to pay a higher service fee for an earlier delivery date, or a lower service fee and a longer waiting time). Their computational experiments showed an increase in agents social welfare. Lai, Sycara, and Li (2008) proposed a protocol that can not only assist agents to make offers efficiently in the n-dimensional space but also give agents sufficient decision flexibility. They showed their efficiency of negotiation outcome by comparing with Pareto equilibrium results. There has been consistent evidence that using an intelligent agent to negotiate with a human counterpart achieves significantly better outcomes than negotiation between two human beings in terms of the social welfare (Kraus, Hoz-Weiss, Wilkenfeld, Andersen, & Pate, 2008; Lin et al., 2008) when the agent is equipped with heuristic algorithms for reasoning and proposal-making. While the results are encouraging, several complexities restrict their significance: (1) Implementation of the computational model. It remains challenging to elicit human preferences in multiple issues. (2) Information exchange mechanism. Computational negotiation agents might not be able to exchange information as efficiently as human beings in occasions where accurate representations are hard to achieve. (3) How "efficient" the negotiation outcome is ultimately depends on human affect and cultural factors, which have not been taken into account by the existing computational researchers.

4 Conclusions

In this chapter, we provided a selective review of two streams of negotiation literature: artificial intelligence and behavioral sciences. We first discuss their key differences and similarities in focus, methodology, and forms of presentation. Then we put the literature in a unified framework under which we summarized how the existing work in both of the two fields treats these four categories. We identify various opportunities to leverage the richness of knowledge in the behavioral literature and the detailed modeling of agent decision making, aiming to consolidate negotiation work that could address more realistic and challenging negotiation situations. At this point, it is useful to suggest some concrete ways in which these two literature can enhance each other. As it has been pointed out earlier in the paper, findings from the behavioral literature have been richer in quantity and complexity. It is thus logical to suggest that the quantitative findings from this literature should be used to inform computational research on negotiation. More specifically, behavioral literature can shed light on how people behave in various bargaining contexts where there is the additional complexity of emotion, stereotyping prejudice and many other similar factors. It is our belief that computational research can use these findings to take negotiation research to the next frontier of capturing uniquely human dimensions, which are particularly relevant in an interaction such as negotiation. One example is the subfield of argumentation-based negotiation (ABN), which so far has been in the way to more realistic negotiation modeling thanks to researchers' efforts in leveraging the work by argumentation theorists (e.g., Amgoud et al., 2007). It is also our contention that this point is especially valuable for culture research, where there is variation across and within culture as well as with regards to the topics being examined.

References

Adair, W. L., & Brett, J. M. (2005). The negotiation dance: Time, culture, and behavioral sequences in negotiation. *Organization Science, 16*(1), 33–51.

Amgoud, L., Dimopoulos, Y., & Moraitis, P. (2007, May). A unified and general framework for argumentation-based negotiation. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems* (p. 158). ACM.

Anderson, N. R., & Neale, M. A. (2005). *All fired up, but who to blame? The benefit of anger and uncertainty in negotiation*. Working paper, Department of Psychology, Stanford University, Stanford, CA.

Anderson, C., & Shirako, A. (2008). Are individuals reputations related to their history of behavior. *Journal of Personality and Social Psychology, 94*, 320–333.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Barry, B., & Oliver, L. (1996). Affect in dyadic negotiation: A model and propositions. *Organizational Behavior and Human Decision Processes, 67*, 127–143.

Balakrishnan, P. V., & Eliashberg, J. (1995). An analytical process model of two-party negotiation. *JehoshuaManagement Science, 41*(2), 226–243.

Braun, P., Brzostowski, J., Kersten G., Kim, J. B., Kowalczyk, R., Strecker, S., et al. (2006). e-Negotiation systems and software agents: Methods, models and applications. In *Intelligent decision-making support systems* (pp. 271–300). London: Springer.

Busch, L-A., & Horstmann, I. J. (2002). The game of negotiations: Ordering issues and implementing agreements. *Games and Economic Behavior, 41*, 169–191.

Carnevale, P. J., & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational Behavior and Human Decision Processes, 37*(1), 1–13.

Curhan, J. R., Elfenbein, H., & Xu, H. (2006). What do people value when they negotiate? Mapping the domain of subjective value in negotiation. *Journal of Personality and Social Psychology, 91*, 493–512.

De Dreu, C. K. (1997). Productive conflict: The importance of conflict management and conflict issue. *Using conflict in organizations*(pp. 9–22). London/Thousand Oaks, CA: Sage.

De Dreu, C., Weingart, L., & Kwon, S. (2000). Influence of social motives on integrative negotiation: A meta-analytic review and test of two theories. *Journal of Personality and Social Psychology, 78*, 889–905.

Deutsch, M. (1949). A theory of competition and cooperation. *Human Relations, 2*, 199–231.

Druckman, D. (1986). Stages, turning points and crises: Negotiating military base rights, Spain and the United States. *The Journal of Conflict Resolution, 30*, 327–360.

Fairfield, K. D., & Allred, K. G. (2007). Skillful inquiry as a means to success in mixed-motive negotiation. *Journal of Applied Social Psychology, 37*(8), 1837–1855.

Faratin, P., Sierra, C., & Jennings, N. R. (2002). Using similarity criteria to make issue trade-offs in automated negotiations. *Artificial Intelligence, 142*, 205–237.

Fogelman-Soulie, F., Munier, B., & Shakun, M. F. (1983). Bivariate negotiations as a problem of stochastic terminal control. *Management Science, 29*(7), 840–855.

Forgas, J. P. (1992). Mood and the perception of unusual people: Affective asymmetry in memory and social judgments. *European Journal of Social Psychology, 22*(6), 531–547.

Forgas, J. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of Personality and Social Psychology, 74*, 129–151.

Frijda, N. H. (1994). Varieties of affect: Emotions and episodes, moods, and sentiments. In *The nature of emotions: Fundamental questions* (pp 197–202). New York: Oxford University Press.

Galinsky, A. D., Seiden, V., Kim, P. H., & Medvec, V. H. (2002). The dissatisfaction of having your first offer accepted: The role of counterfactual thinking in negotiations. *Personality and Social Psychology Bulletin, 28*, 271–283.

Hopmann, P. T., & Walcott, C. (1977). The impact of external stresses and tensions on negotiations. *Negotiations: Social-Psychological Perspectives*. Beverly Hills, CA: Sage.

Kramer, R. M., Pommerenke, P., & Newton, E. (1993). The social context of negotiation effects of social identity and interpersonal accountability on negotiator decision making. *Journal of Conflict Resolution, 37*(4), 633–654.

Kraus, S. (1997). Negotiation and cooperation in multi-agent environments. *Artificial Intelligence, 94*, 79–97.

Kraus, S., Wilkenfeld, J., & Zlotkin, G. (1995). Multiagent negotiation under time constraints. *Artificial Intelligence, 75*, 297–345.

Kraus, S., Hoz-Weiss, P., Wilkenfeld, J., Andersen, D. R., & Pate, A. (2008). Resolving crises through automated bilateral negotiations. *Artificial Intelligence, 172*, 1–18.

Kumar, R. (1997). The role of affect in negotiations. *Journal of Applied Behavioral Science, 33*, 84–100.

Lai, G., Sycara, K., & Li, C. (2008). A decentralized model for automated multi-attribute negotiations with incomplete information and general utility functions. *Journal of Multi Agent and Grid Systems, 4*(1), 39–57.

Lange, A., & Vogt, C. (2003). Cooperation in international environments due to a preference for equity. *Journal of Public Economics, 87*, 2049–2067.

Lewicki, R.J., Bies, R.H., & Sheppard, B.H. (1997). *Research on negotiation in organizations*. Vol. 6. Greenwich, CT: JAI Publishing.

Lin, R., Kraus, S., Wilkenfeld, J., & Barry, J. (2008). Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence, 172*, 823–851.

Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology, 57*, 426–441.

Luo, X., Jennings, N. R., Shadbolt, N., Leung, H., & Lee, J. H. (2003). A fuzzy constraint based model for bilateral, multi-issue negotiations in semi-competitive environments. *Artificial Intelligence, 148*, 53–102.

McGinn, K., & Keros, A. (2002). Improvisation and the logic of exchange in socially embedded transactions. *Administrative Science Quarterly, 47*, 442–473.

Myerson, R. B. (1977). Two-person bargaining problems and comparable utility. *Econometrica, 45*, 1631–1637.

Nash, J. (1953). Two-person cooperative games. *Econometrica, 21*, 128–140.

Olekalns, M., & Weingart, L. R. (2008). Emergent negotiations: Stability and shifts in negotiation dynamics. *Negotiation and Conflict Management Research, 1*(2), 135–160.

Olekalns, M., Brett, J. M., & Weingart, L. R. (2003). Phases, transitions and interruptions: Modeling processes in multi-party negotiations. *International Journal of Conflict Management, 14*(3/4), 191–211.

Pruitt, D., & Rubin, J. (1986). *Social conflict: Escalation, stalemate and settlement*. New York: Random House.

Pruitt, D. G., & Kim, S. H. (2004). Social conflict: Escalation, stalemate, and settlement (3rd ed.). New York: McGraw-Hill.

Raiffa, H. (1982). *The art and science of negotiations*. Cambridge, MA: Belknap/Harvard University Press.

Rangaswamy, A., & Shell, G. R. (1997). Using computers to realize joint gains in negotiations: Toward an "electronic bargaining table". *Management Science, 43*(8), 1147–1163.

Sinaceur, M., & Tiedens, L. Z. (2006). Get mad and get more than even: When and why anger expression is effective in negotiations. *Journal of Experimental Social Psychology, 42*(3), 314–322.

Sycara, K. (1990). Persuasive argumentation in negotiation. *Theory and Decision, 28*(3), 203–242.

Synder, C., & Higgins, R. (1988). Excuses their effective roles in the negotiation of reality. *Psychological Bulletin*, *124*, 23–45.

Thompson, L. (1998). *The mind and heart of the negotiator*. Upper Saddle River, NJ: Prentice Hall.

Van Kleef, G. A., De Dreu, C., & Manstead, A. (2003). *The interpersonal effects of anger and happiness on negotiation behavior and outcomes. IACM 15th annual conference*. Amsterdam: University of Amsterdam.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*, 337–349.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219.

Weingart, L. R., & Olekalns, M. (2004). Communication processes in negotiation: Frequencies, sequences, and phases. *The handbook of negotiation and culture* (pp. 143–157). Stanford, CA: Stanford Business Books.

Weingart, L. R., Hyder, E. B., & Prietula, M. J. (1996). Knowledge matters: The effect of tactical descriptions on negotiation behavior and outcome. *Journal of Personality and Social Psychology, 70*(6), 1205.

Zeng, D., & Sycara, K. (1998). Bayesian learning in negotiation. *International Journal of Human-Computers Studies*, *48*(1), 125–141.

Zetik, D. C., & Stuhlmacher, A. F. (2002). Goal setting and negotiation performance: A meta-analysis. *Group Processes & Intergroup Relations, 5*(1), 35–52.

Zlotkin, G., & Jeffrey, S. (1996). Rosenschein. Compromise in negotiation: Exploring worth functions over states. *Artificial Intelligence, 84*, 151–176.

A Model of Culture in Trading Agents

Gert Jan Hofstede, Catholijn M. Jonker, and Tim Verwaart

Abstract Geert Hofstede's five-dimensional framework is widely used in social sciences and management science to characterize cultures. It has been suggested to build culturally consistent agent characters based on his framework. This chapter stresses the relevance of culture and trust for trade, substantiates why a dimensional model offers a good basis for cultural differentiation of agents, and presents an approach to apply Hofstede's model to develop culturally differentiated agents. The approach is based on knowledge acquisition on a dimension-by-dimension basis and a computational method to integrate the acquired knowledge. The approach has been applied to a multi-agent simulation of a trade game. It is instantiated for the processes of partner selection, negotiation, and the interaction between deceit and trust in trade.

Keywords Cultural dimensions • Trust • Trade • Multi-agent simulation

1 Introduction

Great differences in personality may exist between people. On the other hand, all people around the world have their human nature in common. Human nature includes life in society. One can observe groups of people, on scales ranging from

G.J. Hofstede (✉)
Social Sciences, Wageningen University, Hollandseweg 1, 6706 KN Wageningen, The Netherlands
e-mail: gertjan.hofstede@wur.nl

C.M. Jonker
Interactive Intelligence Group, Delft University of Technology, Mekelweg 4, 2624 PS, Delft
e-mail: c.m.jonker@tudelft.nl

T. Verwaart
LEI Wageningen UR, Alexanderveld 5, 2585 DB den Haag, The Netherlands
e-mail: tim.verwaart@wur.nl

clans to nations, living together and having common behaviors that distinguish them from other groups.

Hofstede (2001, p. 9) defines culture as "*the collective programming of the mind that distinguishes one group or category of people from another*". This implies that culture is not an attribute of individual people, unlike personality. It is an attribute of a group that manifests itself through the behaviors of its members.

A shared understanding of the world is essential for social life. From early childhood on the cultural meaning systems are anchored in the minds of people and permanently reinforced and extended by joint activities with other group members. Kashima (2009) concludes that the joint activities members of a culture engage in "*are in fact the very activities that get their society and culture going. They produce and reproduce the social relationships and cultural resources that enabled these joint activities to begin with*" (Kashima).

Groups share value systems and practices – Hofstede refers to rituals, heroes, and symbols as practices. These are transferred from generation to generation. The practices are the observable characteristics of a culture. Over time and under the influence of contacts with other cultures, practices may change far more easily than the value systems that constitute the core of a culture (Hofstede, 2001).

The present work focuses on the way culture influences trade. The relevance of cultural conflict in trade grows with the current trend of globalization. In a trading situation, culture of the trader manifests itself in four ways. First, culture filters observation. It determines the salience of clues about the acceptability of trade partners and their proposals. Second, culture sets norms for what constitutes an appropriate partner or offer. Third, it sets expectations for the context of the transaction, e.g. the enforceability of regulations and the possible sanctions in case of breach of the rules. Fourth, it sets norms for the kind of action that is appropriate given the other three and, in particular, the difference between the actual and the desired situation.

One of the most prominent, problematic, issues in trade that depend on culture is trust between trade partners (Hofstede, 2006; Hofstede, Spaans, Schepers, Trienekens, & Beulens, 2004). Trust is the aspect of trade where lack of common ground pre-eminently manifests itself. Trust is an important issue in trade, because of the information asymmetries between suppliers and customers, which become more salient in longer supply chains, and because of the inevitable incompleteness of contracts (Williamson, 1985, 1998). Many definitions have been given for trust. For the present work a pragmatic definition is used. Customer's trust in a particular supplier is defined as the customer's estimate of the probability that the supplier will deliver as agreed, even if the supplier has the incentive and the opportunity to defect.

Agent-based modeling offers new opportunities to study trust and culture in trade. Agent-based modeling enables the study of emergent phenomena at an aggregated level that originate from the behaviors and interactions of individuals. As such it provides an analogue to culture: culture has no other representation than in the minds of its members. Culture is a form of collective intelligence, physically represented in the brains of individuals, grounded during the common activities where its observable properties emerge. Trade is such a common activity.

Agent-based models can be applied as a research tool in combination with simulation gaming (Tykhonov, Jonker, Meijer, & Verwaart, 2008). Observations in gaming simulations can lead to theorizing about the decision models of participants. These models can be implemented in the multi-agent simulation, in order to validate that the hypothetical models lead to the behavior that was observed in the gaming simulation.

The context of the agent-based model presented in the present chapter is the TRUST & TRACING (T&T) game. In runs of this gaming simulation, considerable differences in outcomes are observed across cultures. These observations were the reason to develop this model. Meijer (2009) discusses the validity of the gaming simulation, according to validation aspects as defined by Raser (1969): psychological reality, structural validity, process validity, and predictive validity. The T&T game is an abstract game, not representing or predicting the performance of an actual instance of a real world supply chain. It offers process validity in a laboratory environment, based on psychological reality.

This chapter is structured as follows. The first sections to follow are devoted to related work, including the T&T game, the modeling approach, and the research questions. Then Hofstede's dimensions of culture and their hypothetical effects on the processes in the T&T game are discussed. The last three sections present an integrated model of the five dimensions' joint effect, examples of simulation results, and the conclusion of this work.

2   Related Work

There is a wealth of literature on trade and culture that so far has not been considered in formalized models of trade. In agent-based economics, individual traders are modeled as intelligent agents cooperating in an artificial trade environment. The agents are modeled to mimic authentic human behavior as closely as possible. In recent papers the differences between such agents are no longer solely attributed to differences in their individual economic situations. Aspects such as personality and attitude are considered as well, see, e.g. (Jager & Mosler, 2007). Without considering such aspects, the simulations will not correspond to reality. With respect to formalizing the important influence of cultural background on trade, we only found a few papers. These papers study trade at the macro-level, e.g., (Kónya, 2006). That paper presents an equilibrium analysis on the amount countries invest in learning another language and culture and the size and welfare of those countries. Another example is (Bala & Long, 2005). That paper presents a formal model of the influence of trade on culture, i.e., the reverse direction of influence as studied in the current chapter. Other literature also uses macro-level models, such as the gravity model to study the correlation between culture and trade (Guo, 2004).

Kersten (2002) urges the necessity of cultural adaptation of e-Business systems and proposes an architecture that adapts both business logic and user interface. The

rationale for adapting systems to user's cultures is given by Kersten, Köszegi, and Vetschera (2002), who report significant differences in expectations, perception of the opponent, negotiation process, and outcomes of electronic negotiations across cultures. However, no actual implementations of models of culture in e-Business were found. Blanchard and Frasson (2005) report an application of Hofstede's dimensions in a model to adapt e-Learning systems to the user's culture. Recent research on cultural modeling in agents mostly focus on Embodied Conversational Agents (ECA), including non-verbal behavior like facial expressions, gestures, posture, gazing, and silence in conversations (Payr & Trappl, 2004; Rehm, Nishida, André, & Nakano, 2008). CUBE-G (Rehm et al., 2008) is based on Hofstede's dimensions and focuses on modeling into virtual characters the processes of first meeting, negotiation, and interaction in case of status difference.

All models discussed so far have in common that they model culture with the purpose to support human decision making or to improve human-computer interaction. The purpose of the model proposed in the present chapter is to realistically simulate emergent behavior in multi-agent based simulations for research in the social sciences. The aspects of ECA are of less relevance in this context. Agent behavior may be modeled in a more stylized way. An approach that does so for the purpose of multi-agent simulations is that of Silverman et al. (2007) and Silverman, Johns, Cornwell, and O'Brien (2006). They model agents as a composition of biological, personal (personality, culture, emotions), social (relations, trust), and cognitive (decision) modules, completed with modules for perception, memory, and expression. Their approach is a generic structure for modeling the influence of culture on agent behavior – along with factors like stress, emotion, trust, and personality – through Performance Moderator Functions (PMF). It differs from our approach in that it is an environment to implement validated models of culturally differentiated behavior, while our approach aims to develop and validate such models.

The model presented in this chapter is based on the TRUST & TRACING (T&T) game (Meijer, 2009), a research tool designed to study human behavior in commodity supply chains and networks. The focus of study is on trust in stated quality of commodities. The game is played by a group of 12 up to 25 persons that play roles of producers, middlemen, retailers, or consumers. The goal of producers, middlemen and retailers is to maximize profit. The consumers' goal is to maximize satisfaction. Each player receives (artificial) money. Producers receive envelopes representing lots of commodities. Each lot is of a certain type of product and of either low or high quality. High quality products give more satisfaction points than low quality products. A ticket covered in the envelope (so it is not visible) represents quality. The producers know the quality. Other players have to trust the quality statement of their suppliers, or request a product trace at the cost of some money and some damage to the relations with their suppliers. The game leader acts as a tracing agency and can on request and at the cost of a fee determine product quality. In case of deception the game leader will trace transactions and punish deceivers with a fine and public disgrace.

3   Modeling Approach

One must lean on social sciences literature to model culture. Two main streams of research can be distinguished. First, there is the anthropological approach of "thick description", in which specific cultures are studied by detailed and close observation of behaviors during an extensive time-span. Examples are the works of Lévi-Strauss (1992) and Geertz (1973). Second, there is the comparative approach that tries to identify dimensions on which different cultures can be ordered, aiming to develop a classification system in which cultures can be typed by a small number of qualifications. Examples are the models of culture by Hofstede (2001), Schwarz (2009), Trompenaars (1993) and GLOBE (House, Hanges, Javidan, Dorfman, & Gupta, 2004). The approach of that type of research is to characterize cultures by their indices on a limited number of dimensions. The dimensions and the indices of cultures are typically created by factor analyzing massive surveys with standardized questionnaires in many countries. The value of such dimensions largely depends on the questionnaires used in combination with the sets of respondents that are required. Questionnaire studies will be more reliable predictors of behavior if they are about the desired (for self) than if they are about the desirable (for everyone), and also if they are asked to a broad range of types of respondents as opposed to just one type (e.g., students or managers). The resulting models provide a linear ordering of cultures along each dimension, where implicit norms are hypothesized to be stronger or weaker according to the index on the dimension. As authors of dimensional models stress, these same implicit norms carry over to all relationships in society. In all social situations, they act as filters on perception and on action range. This means that there are no specific values for activity x, e.g. 'trade values', in a dimensional model. It also means that a dimensional model is suited for modeling any process that involves social intercourse, including trade and its sub-activities.

Cultural descriptions of the first type provide rich details about values, norms, symbols, beliefs, rituals, social structure, behavioral patterns, etc., in a particular culture. These will prove very useful for facsimile modeling of specific social systems. The model proposed in the present chapter aims to compare the influence of a great diversity of cultures in the standardized environment of a gaming simulation which is by itself an abstraction of social life. For that purpose we need to posit the model at an impartial distance from any single culture. A dimensional model of culture is more suited than a collection of incommensurable rich descriptions. Dimensional models are culture-level abstractions. They do not depict individuals, but average group characteristics, and therefore the agents in our simulation will be iconic for a culture, not specific for any individual.

The work of Hofstede (2001) focuses on differences between national cultures, particularly on differences in value systems across nations. The computational models of the effects of culture proposed in this chapter are based on that work. Although other dimensional models of culture could certainly be used for similar purposes, Hofstede's framework was chosen over possible other candidates (such as Hall, 1976; House et al., 2004; Schwarz, 2009; Trompenaars, 1993) for various

reasons. First, Hofstede's work is parsimonious and accessible, with only five dimensions[1] compared to GLOBE's 18, and with its 1–100 scales. Second, it has a wide scope, compared to Trompenaars', whose dimensions are statistically correlated and can be described as aspects of only individualism and power distance (Smith, Dugan, & Trompenaars, 1996) or Hall who focused on the dimension of individualism (low-context communication) versus collectivism (high-context communication). Those models miss out on issues related to gender roles, anxiety and Confucian values. Third, it has the greatest empirical base of these studies, with a well-matched sample of 117,000 respondents to the original study plus hundreds of replications during a quarter century that validate the model (Kirkman, Lowe, & Gibson, 2006; Schimmack, Oishi, & Diener, 2005). Fourth, it is the most widely used. It has survived fashions and hasty storms of criticism (Smith, 2006; Sóndergaard, 1994). Fifth and most important, it shows continued predictive value for many societal phenomena (Hofstede, 2001; Smith, 2002).

For the endeavor documented in this chapter, not only a model of culture is essential, but also a model of trade. A process model comprising six main processes was elaborated, where possible based on validated models reported in social science or artificial intelligence literature. The six main processes are:

1. trade goal selection (e.g., buy or sell), based on agent role and stock position
2. partner selection, based on the model of Weisbuch, Kirman, and Herreiner (2000)
3. negotiation, based the ABMP architecture (Jonker & Treur, 2001)
4. delivery, truthfully or untruthfully according to results from social psychology
5. acceptance of deliveries, and decision to trust or trace, based on dynamic trust
6. belief updates and trust dynamics (Jonker & Treur, 1999)

Having decided on a model of trade and a dimensional model of culture, the next step was to model the interaction between the two models. The effects of culture were modeled on a dimension-by-dimension basis, using a classical expert systems knowledge acquisition approach: literature study, expert interviews, formalization and verification of the model, and face validation of results. The results of the dimension-by-dimension analysis are described in separate papers per dimension (Hofstede, Jonker, Meijer, & Verwaart, 2006; Hofstede, Jonker, & Verwaart, 2008a; Hofstede, Jonker, & Verwaart, 2008b; Hofstede, Jonker, & Verwaart, 2008c; Hofstede, Jonker, & Verwaart, 2009a). An overview is given in this chapter in Sect. 5.

Subsequently, the models for the individual dimensions were integrated on a process-by-process basis, except for the trade goal selection, which is not modeled to depend on culture. The resulting culturally adapted decision models of the agents are described in separate papers (Hofstede, Jonker, & Verwaart, 2009b, 2010a, 2010b), as is the computational approach to integration of the

---

[1]A sixth dimension 'Indulgence versus Restraint' was added in 2010 (Hofstede, Hofstede & Minkov (2010), after completion of the present model.

dimension-by-dimension expertise (Hofstede, Jonker, & Verwaart, 2011). Section 5.6 presents an overview of the process models and the effects of culture.

The integration into an all-singing, all-dancing model has been verified through a meta-modeling based sensitivity analysis (Burgers, Hofstede, Jonker, & Verwaart, 2010).

The model has been implemented in the CORMAS multi-agent environment.[2]

4   Research Questions

This work aims to differentiate the behavior of artificial agents that simulate aspects of human behavior, as if the agents had a cultural background. Two assumptions underlie this attempt. The first is that it is possible to define models of agents that share universal, globally valid, properties with human behavior at least in the selected processes of trade. The second is that it is possible to formalize expertise about how culture differentiates these models.

To give an example of the first assumption: to a Western mind negotiation may well be perceived as a sequence of alternating explicit proposals, but would another model fit better to practices in other parts of the world? In this research we do not attempt to answer this kind of questions. We assume that social science literature can provide the modeler with a set of theories of human behavior that have universal coverage for the domain of study and that may have parameters that may be different between cultures.

The main hypothesis of this work is related to the second assumption. This hypothesis is that a composition of universal models of human decision making in a particular domain and a dimensional model of culture can simulate believable differentiation of behavior across cultures. With believable we mean that the behavior simulated by the agents can pass face validation by experts in the cultural differentiation in the domain of application.

The main research question "how can a model be formulated that adapts universal decision functions to culture according to a theory about the dimensions of culture" entails the following, more specific questions.

- Given a universal model of human behavior in some domain, can the model of culture's effects be formulated, following an expert systems knowledge acquisition approach?
- Having formulated the knowledge about the effects of culture on model parameters per dimension, how can a believable model of their joint effect be formulated?

---

[2]http://cormas.cirad.fr/indexeng.htm

- How can the sensitivity of such a model be assessed for variation of cultural dimension indices, relational configurations regarding power and group membership, and universal model parameters?
- Can such a model be applied to fulfill its function as a social sciences research instrument in combination with simulation gaming? This function comprises (Meijer & Verwaart, 2005):

  – Validation of models of behavior induced from game observations
  – Testing of hypotheses about dynamics of aggregated results in relation to parameter changes in individual behavior
  – Test design: selection of useful configurations for games with humans

The questions are answered in the context of the T&T game, by applying Hofstede's dimensional theory of culture to three universal domain models, for partner selection, negotiation, and the interaction between deceit and trust.

5 Hofstede's Dimensions and Their Effects on Trade

Hypothetical differences in behavior can be formulated for traders from different cultures on the basis of the works Hofstede (2001) and Hofstede, Hofstede, and Minkov (2010) and observations in the T&T game (Meijer, 2009). While in reality, individual behaviors are co-determined by historical, contextual and personality factors, these can be abstracted away in the present study of trends of behavior in groups of people. Before creating an integrated model, the behavioral tendency associated with each dimension will be reviewed. The dimensions in Hofstede's model, to be discussed in the following subsections, are masculinity versus femininity, individualism versus collectivism, power distance, uncertainty avoidance, and long-term versus short-term orientation. Each of the dimensions is briefly characterized and expected distinctions between the extremes are specified, based on the work of Hofstede and expert consultations.

5.1 Masculinity Versus Femininity

In masculine societies people are expected to place value on measurable performance criteria such as size, speed and quantity. Big is beautiful. Money is good. Rich people are admired. Life is conceptualized as a series of contests and winning is paramount while losing is a disaster. Implicit trust is low; if you get cheated it is your own fault and you are a loser. If you do good, you do it in the large. If you commit crimes, they are large, not petty ones.

Feminine societies are the opposite. Winners are at risk of awakening feelings of jealousy. Small is beautiful, implicit trust is high, and cheaters are looked down upon. Yet small-scale cheating occurs a lot because society is permissive and, in the case of small misdemeanors, forgiving. Penalties are mild. Good intentions are more important than good performance.

Table 1 Expected behavioral distinctions between traders in the T&T game having masculine versus feminine cultural backgrounds

Masculine	Feminine
Deal with anyone	Build relationships
Rapid deals, as many goods as possible	Take time to negotiate, accept small deals
Be a tough negotiator	Show mutual willingness to accommodate
Cheat and expect to be deceived	Trust
Show off with highly valuable goods	Purchase for practical use
Trace and enforce delivery to contract	Avoid notorious deceivers

Sources: Hofstede (2001), and expert consultations

The meaning of trust across cultures is related to the dimension of masculinity versus femininity. In fact, the statement 'Most people can be trusted' was one of the constituents of the dimension in Hofstede's original research. In feminine cultures, people agree with it more. Since then, many others have investigated the variations of the meaning of the concept across cultures. See e.g. Hofstede et al. (2004) for a discussion of the dynamics of trust and transparency across cultures. Hofstede (2006) distinguishes *intrinsic trust* from *enforceable trust*. Intrinsic trust is trust that accepts vulnerability, while enforceable trust is trust in good performance that is backed up by the option of rewarding and punishing the trustee. To sum it up in a simplified way: the former is what people mean by trust in feminine cultures, and the latter is what people mean by trust in masculine cultures.

Table 1 summarizes expected distinctions reported by Hofstede et al. (2006).

5.2 Individualism Versus Collectivism

The variation in basic group size and cohesion between societies has been shown by sociologists, e.g., in the distinction between *Gemeinschaft* (community) and *Gesellschaft* (society) that Tönnies introduced as early as 1887 (Tönnies, 1963). In a *Gemeinschaft*, people share everything, both material and immaterial, whereas in a *Gesellschaft,* private property and other individual-centered institutions are possible. This variation has been confirmed by social-psychological cross-national studies of practices or values. Triandis (1995) and Hofstede (2001) speak of the distinction between *individualism* and *collectivism*. Minkov (2007) showed that the individualist-collectivist continuum is visible in World Values Survey data (he names it universalism versus exclusionism). This dimension has become the main ingredient of theories about cross-cultural business, e.g., in the work of Trompenaars (1993) who posits a number of dimensions of culture that were shown by Smith et al. (1996) to be correlated with individualism versus collectivism. The relevance of this dimension to the management literature lies in the fact that Anglo countries are at the extreme individualistic end of the scale, so business partners from almost any country are bound to have more collectivistic cultures.

Table 2 Expected behavioral distinctions between traders in the T&T game having collectivistic versus individualistic cultural backgrounds

Collectivistic	Individualistic
Maintain harmony, avoid confrontation	Speak your mind
Show favor to in-group customers	Treat all customers equally
No business without a personal relation	Task is more important than a good relation
Relations are given	Build and maintain relations actively
Do not terminate negotiations	Terminate if progress is insufficient
Do not show distrust to in-group members	Show distrust as it is
High relational endowment effect[a]	Low relational endowment effect
Deceive if it does not conflict with group interests	Deceive according to personal morality

Sources: Hofstede (2001), and expert consultations
[a]The effects of deceit, if revealed, and showing distrust by tracing deliveries do more damage to relationships in collectivistic than in individualistic societies

In an individualistic society the various spheres of life, e.g., family, business, and leisure, tend to be separated. To collectivist mindsets, relations are more important than business – and so, business tends to be done among friends and family. The term "in-group" is often used to denote this kind of self-evident unit of social life.

Table 2 summarizes expected distinctions reported by Hofstede et al. (2008b).

5.3 Power Distance

Where the salience of common group memberships for social interaction makes the difference between individualistic and collectivistic societies, the dimension of power distance is about the importance of hierarchy in societies. The dimension runs from egalitarian (*small power distance*, e.g., Anglo, Germanic, and Nordic cultures) to hierarchical (*large power distance*, most other cultures). Hofstede (2001) defines power distance as the extent to which the less powerful accept and expect that power is distributed unequally. In large power distance societies political, economic, and judicial power are interwoven. In small power distance societies a person's power is limited to the purposes of a role and to the situations in which the person plays the role.

Table 3 summarizes expected distinctions reported by Hofstede et al. (2009a).

5.4 Uncertainty Avoidance

Hofstede (2001, p. 161) defines the uncertainty avoidance dimension of culture as "*the extent to which the members of a culture feel threatened by uncertain or unknown situations*". It is important to realize that this has nothing to do with risk avoidance. Uncertainty avoidance is about fear of situations in which "*anything*

Table 3 Expected behavioral distinctions between traders in the T&T game having hierarchical versus egalitarian cultural backgrounds

Large power distance (hierarchical)	Small power distance (egalitarian)
Might is right	No privileges and status symbols
Formal speech; acknowledgement	Talk freely in any context
Dictate, obey	Negotiate
Negotiate only in case of equal power	
Prefer equal status partners	Treat all (potential) business partners equally
Avoid more powerful business partners, but show favor when unavoidable	
Be patient and conceding with powerful partners	
Be careful not to deceive more powerful partners	
Do not show distrust to more powerful partners	
No need to distrust less powerful partners	
Trade products of quality according to status	

Sources: Hofstede (2001), and expert consultations

can happen and one has no idea what" ... "*Uncertainty-avoiding cultures shun ambiguous situations. People in such cultures look for structure in their organizations, institutions, and relationships, which makes events clearly interpretable and predictable. Paradoxically, they are often prepared to engage in risky behavior in order to reduce ambiguities – such as starting a fight with a potential opponent rather than sitting back and waiting*" (Hofstede, p. 148).

People from highly uncertainty avoiding societies or groups do not tolerate ambiguity as to who is a member of their group. They tend to have strict moral criteria as to who fits in: adherents of the same religious subgroup, perhaps, or people from the same region, people with the same profession or status, people who speak their language, people of their gender, or similar clear-cut criteria. Since people from such societies do not easily engage in interactions with others who do not share their most salient group characteristics, trade will often be a within-group activity. Within-group contacts are charged with tokens of loyalty, often through shared ritual that is needed to counteract the stress that people experience.

Societies or groups that are uncertainty tolerant are easy travelers, and will engage in novel activities without needing much time to adjust. Their social interactions tend to be laid-back and they will strike up trade relations with foreigners if the opportunity presents itself.

Table 4 summarizes expected distinctions reported by Hofstede et al. (2008a).

5.5 Long-Term Versus Short-Term Orientation

Behaving as a good, upstanding member of the group is at the core of the lives of all beings that live in social groups (Wilson, 2007). Human beings are intensely

Table 4 Expected behavioral distinctions between traders in the T&T game having uncertainty tolerant versus uncertainty avoiding cultural backgrounds

Uncertainty tolerant	Uncertainty avoiding
Suppression of emotions	Expression of emotions; strong emotions when deceit is revealed
Willingness to take unknown risks	Only known risks are taken
Openness to innovations	Distrust, strong endowment effect
Tolerance of diversity	Avoid strangers
Comfortable with ambiguity and chaos	Need for clarity and structure
Appeal of novelty and convenience	Appeal of purity, high quality
Relaxed negotiation style	Impatient, time is money
Frequency of deceit depends on relationships	Low deceit threshold toward strangers

Sources: Hofstede (2001), and expert consultations

social and they exemplify the point very well. Ensuring the successful functioning of our social groups is a basic requirement for survival. We spend up to 20 years being taught how to act as virtuous members of society. But how to be virtuous? It turns out that different societies have found different answers to that question. The issue of whether the sources of virtue are to be found in the past and present, or in the future, is one of the basic dimensions of culture found by Hofstede (2001). This fifth dimension, called 'long-term versus short-term orientation' by Hofstede, is very important for trade. Its strong correlation with economic growth across 39 countries in the period 1970–2000 testifies to its relevance for trade (Hofstede & Hofstede, 2005, p. 223).

Hofstede (2001, p. 359) gives the following definition of the LTO-dimension: "*Long Term Orientation stands for the fostering of virtues oriented towards future rewards, in particular, perseverance and thrift. Its opposite pole, Short Term Orientation, stands for the fostering of virtues related to the past and the present, in particular, respect for tradition, preservation of 'face', and fulfilling social obligations*". Long-term orientation is correlated with self-effacement. This was found by Minkov (2007) in a meta-analysis of data from the World Values Survey. The individual thinks of itself as a small element within the continuity of life. Hence, learning and developing one's capacities is more important than winning a particular game, or obtaining a particular result. In contrast, short-term orientation correlates with self-enhancing values in which being successful in a game is a desirable thing that will improve one's reputation.

Table 5 summarizes expected distinctions reported by Hofstede et al. (2008c).

5.6   *The Joint Effects of Hofstede's Dimensions*

The preceding subsections describe how agent behavior in trade is expected to vary across cultures along a single dimension. This approach enables knowledge acquisition. An expert can explain the effects of individual dimensions, but cannot oversee the consequences of varying the indices simultaneously. As was shown

Table 5 Expected behavioral distinctions between traders in the T&T game having short-term versus long-term oriented cultural backgrounds

Short-term oriented	Long-term oriented
Immediate gratification of needs expected	Deferred gratification of needs accepted
Short-term virtues taught: social consumption	Long-term virtues taught: frugality, perseverance
Spending, top quality demanded, impatient	Saving, investing, patient
The transaction bottom line, quick profits	Building a strong market position
Show off, by trading with high status partners	Invest in long-term relations

Sources: Hofstede (2001), and expert consultations

by Hofstede et al. (2006, 2008a, 2008b, 2008c, 2009a), the expertise laid down in narrative form can be formalized into production rules that work in two ways: modify (i.e. either increase or decrease) the values of parameters in decision functions and modify the relevance of some decision functions.

The formal rules can be implemented in agents. In multi-agent simulations results can be generated for different values of the particular cultural index and for different relational configurations (status differences, different in-groups, differences in initial trust etc.). These results can be used for verification of the implementation of the rules, for face validation of results, and even for educational purposes. However, to simulate actual national cultures, all dimensions must be taken into account simultaneously. A computational approach to this integration has been developed (Hofstede et al., 2011). This approach, like the models of the individual dimensions, is based on the adaptation of default parameter values in decision models. The adaptation is made on the basis of the values of "cultural factors". Some dimensions adapt the perceived relevance of relational attributes, such as group membership (in-group versus out-group), status difference, and trust. Cultural factors combine dimension scores and relational attributes (see Table 10 at the end of this chapter). The integrated effect of culture on agent behavior can be modeled as a function h that maps a vector of cultural factors $\vec{f}$ and a vector of default values of model parameters $\vec{x}$ to a vector of culturally adjusted parameters $\vec{x}'$:

$$h\left(\vec{f}, \vec{x}\right) = \vec{x}'. \tag{1}$$

The hypothesis of this work entails that, given the set of decision functions, a dimensional theory of culture can be used (a) to identify the cultural factors to be taken into account and (b) to define the mapping h. If this is possible, the agent modeling can benefit from vast bodies of social sciences literature that describe the differentiation of many behaviors along the dimensions of the cultural model.

The computational integration approach is based on the following assumptions.

1. Given that dimensional models of culture aim to provide for each dimension a linear ordering of the strength of phenomena associated with that dimension, the effect of each cultural factor may be modeled as a strictly monotonic function r_{ijk} that adapts the i-th parameter to the k-th factor associated with the j-th dimension.

r_{ijk} can be seen as a member of a set of functions that can be indexed by the labels of cultural factors and parameters as arguments.

2. As long as there is no further evidence, a first order approach can be taken, i.e., let r_{ijk} adjust x_i proportionally to f_{jk} from its default value in the direction of the extreme values $\varepsilon_{ijk}^+ > x_i$ and $\varepsilon_{ijk}^- < x_i$, with ε_{ijk}^+ and ε_{ijk}^- universal, i.e. not dependent on culture, for a particular domain.

3. The interaction between decision function parameters does not depend on culture. This assumes that a decision model can be formulated in such a way that any parameter can be modified for culture without taking the values of the other parameters into account. This is not a very restrictive assumption. For instance if a decision function $\Delta(x_i, x_{i'})$ requires parameters $x_i < x_{i'}$, the parameters cannot independently be modified, but they can after substitution: $\Delta(x_i, x_{i'}) = \Delta(x_i, x_i + x_{i''}) = \Delta'(x_i, x_{i''})$ with $x_{i''} > 0$.

4. The joint decreasing and the joint increasing effect of cultural dimensions can compensate for each other. This expertise is confirmed by expert statements, e.g. (in cultures with high power distance) "*The powerful dictate the conditions. The less powerful have to accept. In feminine or collectivist cultures the powerful may exercise restraint, ...*".

5. For the increasing and for the decreasing effects, the effect with maximal influence is dominant: effects in the same direction do not reinforce each other. According to expert knowledge, if several factors influence a parameter in equal direction, it is sufficient for one to be maximal in order to sort maximal effect (disjunctive factor influence, see e.g. "*feminine or collectivist*" under 4. above).

6. Cultural factors working in the same direction do not reinforce each other. For instance, in Table 2 three factors are identified to have increasing effect on deceit threshold d_b. If two of the factors have effect 0.5 and one has effect 0.2, their joint effect is 0.5; not 0.4 (the average) or another linear combination (see 5. above); not 0.8 (probabilistic) or another product combination.

Under these assumptions, the mapping h can be written as a set of functions g_i, one function for each parameter (Hofstede et al., 2011):

$$g_i\left(\vec{f}, x_i\right) = x_i + \left(\hat{\varepsilon}_i^+ - x_i\right) \max\{f_{jk}|l_{jk} \in L_i^+\} + \left(\hat{\varepsilon}_i^- - x_i\right) \max\{f_{jk}|l_{jk} \in L_i^-\}. \tag{2}$$

The inputs to this equation are the (domain-dependent) default, minimal and maximal values of the parameter, and the actual values f_{jk} of the cultural factors labeled l_{jk}. The cultural factors are identified in the dimension-by-dimension knowledge acquisition. L_i^+ stands for the set of cultural factors indicated to have increasing effect on x_i and L_i^- stands for the set indicated to have decreasing effect.

In order to apply this approach, one needs models for the decisions that the agents make in the simulated processes. The following subsections propose models for the trade processes of partner selection and negotiation in the pre-contract phase of transactions, and deceit and trust in the post-contract or delivery phase.

5.7 Partner Selection

The decision model for partner selection (Hofstede et al., 2009b) is based on the reinforcement learning of expected utility proposed by Weisbuch et al. (2000). According to the model of Weisbuch et al., agents select their business partners at random, with probability:

$$P_b = \frac{\exp(\beta J_b)}{\sum_{b'} \exp(\beta J_{b'})}, \tag{3}$$

where J_b represents the preference for business partner b, based on the agent's belief about the partner's fairness (defined as experience of utility of previous deals) and affected by the agent's culture and relationship with the partner. The effects of culture and relationship are computed according to Table 10 (at the end of this chapter). If the value of β is high, the agent has a tendency to be loyal to partners with which it has successfully dealt before. If the value of β is low, the agent is likely to display shopping behavior, frequently trying new partners. Weisbuch et al. have validated this model in their research at the Marseille wholesale fish market, where both types of agent behavior prevail: some agents are loyal and some consistently show shopping behavior.

Based on the descriptions of individual dimensions given in the preceding section, the value of β, representing loyalty, is expected to depend on culture: increased to a maximal value in long-term oriented societies, and decreased to a minimal value in uncertainty-avoiding or masculine societies.

For each partner, the agent maintains a belief J'_b about the partner's fairness:

$$J'_b(n) = (1 - \gamma) J'_b(n - 1) + \gamma U_b(n), \tag{4}$$

where γ represent a learning parameter and $U_b(n)$ the utility of the n-th negotiation result with j; $U_b(n) = 0$ if the negotiation was terminated without agreement. Based on the descriptions given in the preceding section, the value of γ is expected to depend on culture: increased in feminine, decreased in uncertainty avoiding cultures.

If an agent has no negotiation going on, it checks for recently received proposals. It may have recent proposals from several agents simultaneously. From the simultaneous proposers, it selects the agent z that has the maximal acceptability A_z. of all proposers and subsequently decides whether to accept the proposal to negotiate with z or to start new partner search, and propose to a partner selected according to Eq. 3. This decision is modeled as a Bernoulli variable:

$$p \text{ (start negotiation with } z) = A_z; \tag{5}$$

$$p \text{ (start new partner selection)} = 1 - A_z. \tag{6}$$

Acceptability A_b of an agent b is set equal to J_b, but for agents from hierarchical societies the value is modified to express that agents having that cultural background are inclined to accept a higher-ranked proposer even if they do not prefer the partner.

Table 10 (at the end of this chapter) summarizes the cultural factors taken into account to modify the parameters and the variables of the partner selection process, applying Eq. 2 and using appropriate settings for the default, minimal, and maximal values of these parameters and variables.

5.8   Negotiation

The negotiation process (Hofstede et al., 2010a) is modeled according to the ABMP architecture (Jonker & Treur, 2001). The similarity of ABMP to human negotiations has been validated (Bosse, Jonker, & Treur, 2004). The ABMP process is an exchange of bids, starting with a bid by one of the partners. The agents evaluate bids using a utility function.

The agent model uses the utility function proposed by Tykhonov et al. (2008):

$$U = w_{value} V + w_{qual} Q + w_{risk} R, \tag{7}$$

with $0 \leq w_{value} \leq 1$, $0 \leq w_{qual} \leq 1$, $0 \leq w_{risk} \leq 1$, and $w_{value} + w_{qual} + w_{risk} = 1$. The terms of this function represent the economic value of the transaction and the effects of agent's quality preference and risk avoidance, respectively. For a further specification of the computation of the utility function's terms, we refer to Hofstede et al. (2010a).

ABMP is a multi-attribute concession strategy. An agent prepares a bid that is a utility concession to its previous bid. In the present simulation, the agent prepares the first bid by composing an attribute configuration according to its quality preference and risk attitude. After the price is set according to the agent's belief about the market price range for the product (a seller sets the price at the maximal value, a buyer at the minimal value), the utility U_1 of the first bid to be made is computed. In subsequent bids the partners make concessions with respect to the utility of their bid.

Concession factor η and negotiation speed χ are the parameters that govern the concession making. *Concession factor* η, $0 < \eta < 1$, is the fraction of the opening bid's utility that the agent is willing to give in during the negotiation. It determines the minimum utility that is acceptable to an agent, also called the reservation value:

$$U_{\min} = (1 - \eta) U_1, \tag{8}$$

where $U_{\min}$ represent the reservation value, and U_1 the opening bid's utility.

Negotiation speed χ, $0 < \chi < 1$, is the fraction of difference between the agent's previous bid and the minimum utility that an agent uses to determine the target utility of its next bid:

$$U_{n+1} = U_{\min} + (1 - \chi)(U_n - U_m i n), \qquad (9)$$

where $U_{\min}$ represent the reservation value, and U_1 the opening bid's utility. In the current simulation a bid is composed by changing the attribute values at random and setting the price to yield the target utility.

After calculation of the utility of a partner's bid and the target utility of its own next bid, the agent decides whether to accept partners bid or not, governed by the utility gap parameter ω.

Acceptable utility gap ω is the maximal difference between own target utility and last partner's bid's utility for which an agent will accept partner's bid.

If the target utility minus the partner's last bid's utility is greater than the acceptable utility gap, the agent does not accept and has to decide about its next action. It can terminate the negotiation for several reasons. First, partner's bid may be interpreted as unrealistic if its utility is too far below the minimum utility. Second, an agent may be dissatisfied by the progress in partner's bids. Third, there may be no more room for a substantial change of attributes to make a bid with the target utility. In the latter case the agent terminates the negotiation. In the first two cases the probability that the agent terminates the negotiation depends on the impatience parameter ι, $0 < \iota < 1$. The agent terminates the negotiation with probability ι if partners bid is considered unrealistic, i.e. less than $(1-\iota)U_{\min}$, or partner makes insufficient concessions, i.e. less than $(1-U_{\min})\iota$ over the last three rounds.

The weight factors in the utility function and the ABMP parameters depend on the negotiation domain and on culture. Table 10 (at the end of this chapter) summarizes the effects of culture for the negotiation process. Equation 2 can be applied to modify the values of weights and parameters, given default, minimal and maximal values for a domain, e.g. the purchase of a supercomputer by a university or transactions in a wholesale food market.

5.9 Opportunism and Trust

For a simulation of human trust and deceit, a strictly rational model is not sufficient (Hofstede et al., 2010b). In intelligent agent research, much attention has been paid to trust. Little research has been published about the simulation of deceit. Some authors modeled deceit as a rational strategy to gain advantage in competitive situations (Castelfranchi, Falcone, & De Rosis, 2001; Ward & Hexmoor, 2003). A strictly rational approach of deceit neglects the emotional impact that deceit has, not only on the deceived, but also on the deceivers. Feelings of guilt and shame result from deceiving and the extent to which these feelings prevail differs across cultures (Triandis et al., 2001). People have emotional thresholds for deceit that cannot be explained from rational evaluation of cost and benefit, but that are based on morality and cooperative attitudes (Boles, Croson, & Murnighan, 2000; Steinel & De Dreu, 2004; Wirtz & Kum, 2004). Once deceived, people react to an extent

that goes beyond rationality (Boles et al., 2000), especially when they are pro-social rather than selfish (Steinel & De Dreu, 2004). In human decision making a model based on *fair trade* prevails over a model of *opportunistic betrayal* (Olekalns & Smith, 2009). In addition to psychological factors, rational economic motives can be given for the human inclination to cooperative behavior (Hwang & Burgers, 1999).

The decision to deceive depends on opportunity, motive, and attitude. A supplier has an opportunity to deceive if a high quality product has to be delivered and the customer did not require a certificate. A supplier has a motive to deceive if an extra profit can be gained by deceiving, for instance if the customer negotiated a guarantee and the supplier expects the customer to trust the delivery. Whether a supplier actually deceives when opportunity and motive are present, depends on the supplier's threshold toward deceit. The threshold depends on the supplier's morality and may be reinforced by recent penalties or the relationship with the customer. For agents in the TRUST & TRACING game, Hofstede et al. (2010b) model the decision to deceive as a Bernoulli variable with probability of deceit

$$p \text{ (deceit)} = q \, (1 - c) \, m_b \, (1 - d_b), \tag{10}$$

where q represents the quality agreed in the current contract ($q = 1$ for high quality; $q = 0$ for low quality or no opportunity); $c = 1$ if certification has been agreed (no opportunity); $c = 0$ otherwise; m_b represents the supplier's motive or rationale to deceive customer b ($m_b = 1$ if the supplier expects an extra profit from deceit; $m_b = 0$ otherwise); d_b represents on the interval [0, 1] seller's threshold for deceit toward customer b, where $d_b = 1$ represents perfect truthfulness. d_b is influenced by seller's personal traits and values (like risk aversion and morality), power and group relations, and seller's estimate of customer's benevolence, i.e., seller's trust that the customer will accept deliveries without tracing. A more detailed description of the deceit model is given by Hofstede et al. (2010b).

Trust and distrust develop during social interactions. The only sources of information that can be taken into account in the simulation of the Trust and Tracing game are negotiation outcomes and tracing reports, which are relevant in reality as well. Every successful negotiation resulting in a transaction will strengthen partners' trust in each other. However, customers can decide to trace a delivery and this can have its effects on mutual trust. First, if tracing reveals deceit, the customer's trust in the seller will be reduced. Second, the fine and the reputational damage resulting from revealed deceit will reinforce the supplier's honesty. However, reinforced honesty will decay in the course of time. Third, the supplier delivering truthfully may be offended by tracing. To maintain a good relation, customers may exercise restraint to trace. Tracing will always reduce the supplier's belief about customer's benevolence. The following dynamics have to be modeled:

- development of trust and benevolence belief by successful negotiations;
- for customers: reduction of trust in case of revealed deceit;
- for suppliers: reinforcement of honesty in case of revealed deceit;

- for suppliers: decay of reinforced honesty to a base level;
- for suppliers: reduction of benevolence belief in case of tracing.

Jonker and Treur (1999) develop a classification of trust dynamics. The most realistic type for trading situations is slow positive – fast negative: it takes a series of positive experiences to develop trust, but trust can be destroyed by a single betrayal. A consumer's trust in supplier b after the n-th experience is updated as:

$$
\begin{aligned}
t_b(n) &= t_b(n-1) + u^+(1 - t_b(n-1)) &&\text{if } n^{\text{th}} \text{experience is positive,} \\
t_b(n) &= t_b(n-1) - u^- t_b(n-1) &&\text{if } n^{\text{th}} \text{experience is negative,} \\
t_b(n) &= t_b(n-1) &&\text{if } n^{\text{th}} \text{experience is neither positive} \\
& && \qquad \text{nor negative,} \qquad\qquad (11)
\end{aligned}
$$

with $0 < u^+ < u^- < 1$, where $t_b = 1$ represents complete trust and $t_b = 0$ represents complete distrust; a successful negotiation counts as a positive experience; a tracing report revealing deceit counts as negative; all other experiences are considered neither negative nor positive with respect to trust.

A supplier's belief about a customer's benevolence is updated by the same mechanism. A successful negotiation counts as a positive experience. Tracing counts as a negative experience for a supplier, whether it reveals deceit or not. An additional effect of revealed deceit on the supplier's part is that supplier's current honesty $H(n)$ (a personal trait, representing the inclination to deliver truthfully) is reinforced to 1, representing maximal honesty. H will subsequently decay to a base value H_{min} on each interaction, whether it is successful or not, with a decay factor φ.

$$
H(n) = H_{min} + \varphi(H(n-1) - H_{min}), \quad \text{with } 0 < H_{min} < 1 \quad \text{and} \quad 0 < \varphi < 1.
\tag{12}
$$

Trust developed according to Eq. 11 represents an internal state of mind, a belief about a partner. It can be seen as a subjective probability that a partner will not defect even if motive and opportunity are present. According to Castelfranchi and Falcone (1998) trust exists as a state of mind, as a decision, and as a social behavior. Because of the repercussions distrust may have as a social behavior, the decision to trust does not necessarily correspond to trust as a state of mind. The inclination to trust depends on cultural factors. The decision to trust or to trace is modeled in the simulation as a Bernoulli variable, with

$$
p(\text{trace}) = T', \tag{13}
$$

$$
p(\text{trace}) = 1 - T', \tag{14}
$$

where the culturally adapted inclination to trace T' is computed by applying Eq. 2 on the base value of $T = q(1-c)(1-t_b)$, using Table 10. Table 10 (at the end of this chapter) summarizes culture's effects on the variables related to trust, deceit, and tracing.

6 Simulation Examples

This section presents some simulation results. The first example confirms that the model is sensitive to culture and offers face validation at extreme values of Hofstede's indices. The second example illustrates that model outputs are more sensitive to common group membership for agents configured with Japanese settings than for USA settings. Then, sensitivity analysis results are presented to confirm that in different cultures, different rules for agent behavior are relevant. The last example validates the feasibility of the model as a research instrument to be used in combination with gaming simulation, aiming to improve models of agent behavior. It uses results of games played in the Netherlands and the USA.

The simulation results should not be interpreted as to give exact quantitative values. The results illustrate stylized facts, such as "common group membership is more relevant in Japan than it is in the USA".

Table 6 presents aggregated outputs of games simulated in an environment where the agents can select a trade partner, negotiate, deceive or deliver truthfully, and update their beliefs. The results show that the ease with which agents reach an agreement, the frequency of negotiation failure, and the quality level depend on culture in a believable way. Table 7 displays the parameter values used.

Much of the literature on cultural difference focuses on individualism versus collectivism and the associated phenomena such as direct versus indirect communication styles. Common group membership determines social relations in

Table 6 Average results of simulated negotiations for cultural stereotypes, with the value for the particular dimension set to either 0.1 or 0.9 and the values for the other dimensions set to 0.5

Culture type	Conditions	Number of transactions	% Failed negotiations	% Top quality
Large power distance	All high status	44	57	97
	All low status	50	60	0
	Customer higher	77	45	98
	Supplier higher	4	92	0
Small power distance		72	49	2
Uncertainty avoiding	Similar partners	29	71	76
	Different partners	27	73	87
Uncertainty tolerant		49	58	1
Individualistic		66	50	1
Collectivistic	In-group partners	117	13	61
	Out-group partners	39	65	0
Masculine		36	71	80
Feminine		61	45	0
Long-term oriented		55	52	0
Short-term oriented	Equal status	24	72	95
	Diverse status	57	47	91

Data source: Hofstede et al. (2010a)
8 suppliers; 8 customers; 10 runs of 200 time steps for each configuration; parameter values as in Table 7

Table 7 Parameter values
used in the simulation runs

Type of value	$w_q{}'$	$w_r{}'$	χ	η	ω	ι
Default value	0.1	0.1	0.7	0.2	0.02	0.3
Maximal value	0.5	0.5	1	0.5	0.1	1.0
Minimal value	0	0	0	0	0	0.1

$w_q{}'$ quality preference, $w_r{}'$ risk aversion, χ concession
factor, η negotiation speed, ω acceptable utility gap, ι
impatience
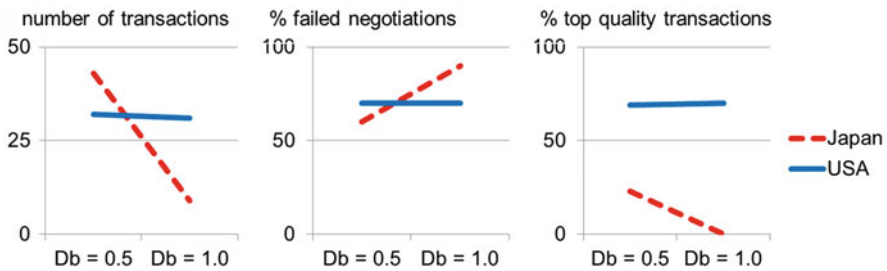
Fig. 1 Simulation results with different settings of group distance between suppliers and customers (Data source: Hofstede et al., 2010a)
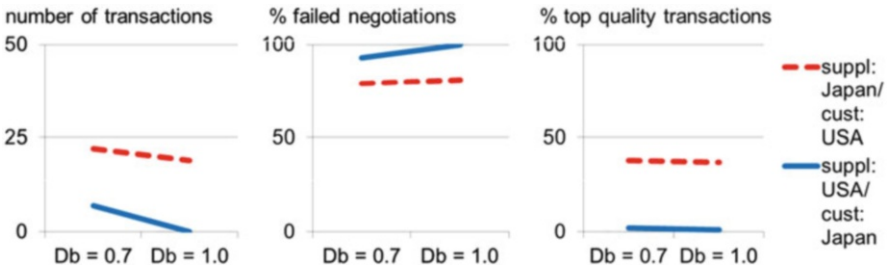
Fig. 2 Simulation results with different settings of group distance between Japanese suppliers and USA customers vice versa (Data source: Hofstede et al., 2010a)

collectivistic societies. Group membership is known to be very relevant in Japan (e.g., Hall, 1976). The cultures of Japan and the USA differ considerably on individualism versus collectivism, but also on other indices (Hofstede, 2001):

– Japan: $PDI^* = 0.54$, $IDV^* = 0.46$, $MAS^* = 0.95$, $UAI^* = 0.92$, $LTO^* = 0.80$;
– USA: $PDI^* = 0.40$, $IDV^* = 0.91$, $MAS^* = 0.62$, $UAI^* = 0.46$, $LTO^* = 0.29$.

It is interesting to see if the influence of group membership emerges in simulations, in spite of the differences on other dimensions. Figure 1 presents simulation results for Japan and USA, with parameter settings as in Table 7. For the "Japanese" agents, group membership considerably influences the success of negotiations and the trust required for trading high quality products.

Trade may run less smoothly between different cultures. Figure 2 presents simulated results of trade between Japanese and USA agents. The results show that for USA traders in Japan it pays to invest in personal relations in order to reduce

group distance. Overcoming distrust is hard for USA traders acting as suppliers to Japanese agents. This phenomenon is not associated with the difference on the IDV index, but with UAI and LTO.

The example of Japanese versus USA traders indicates that the salience of parameters may be different across cultures, in combination with trade roles. Burgers et al. (2010) present results of Monte Carlo based sensitivity analyses for the model proposed in the present chapter. Different outputs (e.g., number of transactions, number of certificates) are sensitive to different parameters. Table 8 displays contributions to variance of the number of transactions in a game. The table presents a mean value and shows that considerable differences between countries prevail. It should be noted that the values presented in Table 8 have no actual meaning for simulations results, but they indicate that it depends on culture which parameters should be given most attention to calibrate the model.

The remaining part of this section presents an example of a tour through the research cycle where a model based on theory is implemented in the multi-agent simulation, multi-agent simulation data is validated against gaming simulation observations, validation fails, theory and the model are adapted, the multi-agent simulation is adapted, and new simulation outputs are validated against the observations and are found to support the adapted theory. The example supports the validity of the multi-agent simulation as a research instrument in combination with gaming simulation. The example has been reported by Hofstede et al. (2010b).

The following hypotheses, to be tested against simulation results, can be formulated from observations in the TRUST & TRACING game (Meijer, 2009).

1. Average quality of products traded in games played in the USA is higher than in the Netherlands.
2. Certification ratio (i.e., the fraction of high quality transactions where a certificate was negotiated) is higher in the USA than in the Netherlands.
3. Defection ratio (i.e., the frequency of deceit in uncertified high quality deliveries) is higher in the USA than in the Netherlands.
4. Tracing ratio (i.e., the fraction of uncertified high quality deliveries for which a trace is requested after delivery) is higher in the USA than in the Netherlands.

To test these hypotheses, 310 simulations were run, with cultural indices for the USA and the Netherlands taken from Hofstede (2001):

– Netherlands: $PDI^* = 0.38$, $IDV^* = 0.80$, $MAS^* = 0.14$, $UAI^* = 0.53$, $LTO^* = 0.44$;
– USA: $PDI^* = 0.40$, $IDV^* = 0.91$, $MAS^* = 0.62$, $UAI^* = 0.46$, $LTO^* = 0.29$.

The other agent parameters are randomly generated for each run pair.

Analysis of the results confirmed hypotheses 1, 3, and 4, but did not confirm hypothesis 2. The certification ratio as defined in hypothesis 2 was approximately equal for the USA and The Netherlands. In these simulations the negotiation model reported in Hofstede et al. (2010a) is applied. According to that model, customers do not take a differentiation of certification cost between themselves and suppliers into account. This difference was found to be an important factor

Table 8 Mean top marginal variance values (of 62 countries) and data for the countries that have the maximum TMV score for a parameter

National culture	Group distance	Mean status	Initial trust	Partner pref.	Conces. factor	Negot. speed	Quality pref.	Risk avoid
Mean (n = 62)	4.1	1.0	3.4	1.8	24.8	30.2	0.5	1.6
Indonesia	16.9	0.1	0.2	0.0	11.6	40.9	0.0	0.0
Morocco	0.7	8.7	3.3	4.7	17.5	37.9	0.0	0.0
Hungary	0.0	0.0	11.5	1.9	37.6	1.4	2.4	11.5
Uruguay	4.9	1.7	8.5	5.5	23.4	24.1	0.0	0.0
Netherlands	0.7	0.0	1.6	0.3	46.2	28.8	0.0	2.1
Iran	1.3	3.1	0.9	0.4	10.3	56.2	0.6	0.0
Austria	0.0	0.0	3.8	2.6	27.1	11.5	4.8	6.9
Japan	0.5	0.0	8.3	1.6	30.8	0.9	0.1	15.8

Source: Burgers et al. (2010)

Table 9 Test data for 310 run pairs for USA and NL

Average of 310 runs	USA	NL	Test stat.[a]	Sample[a]	Probability[a]
Number of transactions	72	61	219	302	< 0.001
Quality ratio	0.37	0.15	277	285	< 0.001
Certification ratio	0.48	0.41	191	281	< 0.001
Defection ratio	0.25	0.13	128	154	< 0.001
Tracing ratio	0.40	0.07	169	177	< 0.001

Source: Hofstede et al. (2010b)
[a]Test statistic, effective sample size, and two-sided probability level for Sign test

in the gaming simulations in the USA. The negotiation model was modified to take tracing fee differences and probability to trace into account. The original equation for customers' risk in an uncertified deal is:

$$r_{customer} = (1 - c)(1 - t_b)q. \tag{15}$$

The modified equation is

$$r_{customer} = (1 - c)(1 - t_b)q + T'\phi_{customer}, \tag{16}$$

with $\phi_{customer}$ representing customers' tracing fee.

The simulation was repeated after replacing Eq. 15 with Eq. 16 and setting the tracing fee equal to 0.2 for suppliers and 0.3 for customers. Table 9 summarizes the results. The simulation results confirm hypotheses 1 through 4. Differences are found to be significant for all variables, with $p < 0.001$ according to the Sign test. This example illustrates how the combination of multi-agent simulation and gaming simulation can improve models of culturally differentiated agent behavior.

7 Conclusion

De Rosis, Pelachaud, & Poggi (2004) suggested to apply Hofstede's theory to build culturally consistent agent characters. The research presented in this chapter shows how this can be done. It shows that rules for cultural differentiation of universal models of aspects of human behavior can be formulated, following an expert systems knowledge acquisition approach. Working on a dimension-by-dimension basis reduces the complexity to a level where a domain expert can specify the rules and verify the results. Doing this for more dimensions simultaneously is too complex for an expert. Furthermore, this chapter proposes an approach to integrate the rules for individual dimensions into a joint effect, resulting in believable cultural differentiation of agent behavior. This approach has been applied to three domains in trade: partner selection, negotiation, and the interaction between trust and deceit. The resulting agent models have been applied in a multi-agent simulation.

A meta-modeling approach for sensitivity analysis of the simulation model has been developed (Burgers et al., 2010). Sensitivity analysis indicates that the outcomes and sensitivity for parameters in the universal models strongly depend on the cultural settings. It is found to be necessary to perform sensitivity analysis to statistics at the aggregated level as well to statistics of individual agent performance. However, the method for sensitivity analysis at the individual level needs further development (Burgers et al.).

Face validity of the results for the extreme cultural situations – called "Synthetic Cultures" by Hofstede and Pedersen (1999) – indicates that the approach proposed in the present chapter is feasible for the development of agents which can be used for training purposes and educational simulations. Application of this model in, for instance, affective agents as it has been suggested by De Rosis et al. (2004) would require further validation.

A model as described in this chapter is what Gilbert (2008) has called a middle range model. Gilbert asserts that the generic nature of such a model means that it is not possible to compare its behavior with any observable instance in the real world. Having found some instance where the model reproduces stylized facts that resemble real world or gaming observations, cannot serve as a general validation and does not guarantee the correctness of predictions made for other cultural situations. Validation of this type of model is an ongoing process. Confidence that the model can predict the effects of culture in new situations grows according to the number of cases where its outputs correspond with those of gaming simulations or other data.

The pragmatic validation of the model – for its use in combination with gaming simulation as a research tool for the social sciences – is also an ongoing process. The example given in the preceding section of this chapter supports the usability of the model for hypothesis testing and positively answers the research question "Can such a model be applied to fulfill its function as a social sciences research instrument in combination with simulation gaming?". In cases where the model does not correspond with data from games or other sources, these results can be used to improve the model and support the advancement of science.

Culture research is an ongoing process as well. The first version of Hofstede's model was defined in four dimensions. The dimension of long-term versus short-term orientation was discovered in a separate research program carried out in China, with questions devised by people with Asian mindsets. Hofstede's model has recently been extended with the dimension of "*Indulgence versus Restraint*" found by Minkov (2007) using World Value Survey data, and included in Hofstede et al. (2010). To include this new dimension into the agent model would require the acquisition and formalization of expert knowledge, as it has been performed for the other dimensions, eventually resulting in the representation of extra cultural factors and their effects in Table 10 (appendix).

Extensions of the width of modeled behavior are also to be considered, and might increase the simulation's validity. For instance, an important component not included in the present work but possibly very influential in real trade networks is contagion of behavior through imitation and social learning by means of a reputation mechanism.

Appendix

Table 10 Increasing (+) and decreasing (−) effects of relevant cultural factors per decision function variable

Dimension index	Culture and relational characteristics	Cultural factor[a]	Partner selection				Negotiation						Deceit and trust			
			β	γ	J_b	A_b	w_q	w_r	η	χ	ω	ι	d_b	T	$u-$	$u+$
PDI	Large power distance	PDI^*														−
	Self status high	$PDI^* s_a$					+									
	Self status low	$PDI^*(1-s_a)$					−									
	With higher ranked partner	$\max\{0,PDI^*(s_b-s_a)\}$			−	+		+	+		+	−	+	−		
	With lower ranked partner	$\max\{0,PDI^*(s_a-s_b)\}$			−			−						−		
UAI	Uncertainty avoiding	UAI^*		−			+			+		+		+	+	
	With stranger	$UAI^* \cdot D_b$						+					−	+		
	Uncertainty tolerant	$1-UAI^*$	−													
IDV	Individualistic	IDV^*														
	With trusted partner	$IDV^* t_b$			+											
	Collectivistic	$(1-IDV^*)$													+	
	With in-group partner	$(1-IDV^*)(1-D_b)$			+				+							
	With out-group partner	$(1-IDV^*)D_b$								−		−	−	−		
MAS	Masculine (competitive)	MAS^*	−				+	+		+		+	−	+	−	
	Feminine (cooperative)	$1-MAS^*$		+	+		−			−		−	−	−		
	With trusted partner	$(1-MAS^*)t_b$			+											
LTO	Long-term oriented	LTO^*	+				−	−					+	−	+	
	With trusted partner	$LTO^* t_b$			+											
	Short-term oriented	$(1-LTO^*)$					+									
	With well-respected partners	$(1-LTO^*)s_b$			+				+				+			
	With other partners	$(1-LTO^*)(1-s_b)$											−			

Hofstede et al. (2009b, 2010a, 2010b)

[a]PDI^*, UAI^*, IDV^*, MAS^*, and LTO^* represent Hofstede's indices;s_a the agent's status, s_b partner's status, D_b group distance, t_b trust; all in [0,1]

References

Bala, V., & Long, N. V. (2005). International trade and cultural diversity with preference selection. *European Journal of Political Economy, 21*, 143–162.

Blanchard, E. G. M., & Frasson, C. (2005). Making intelligent tutoring systems culturally aware. In *The International Conference on Artificial Intelligence, ICAI'05*, Vol. 2. pp. 644–649. Las Vegas, Nevada: CSREA Press.

Boles, T. L., Croson, R. T. A., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes, 83*, 235–259.

Burgers, S. L. G. E., Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2010). Sensitivity analysis of an agent-based model of culture's consequences for trade. In M. LiCalzi et al. (Eds.), *Progress in artificial economics. Lecture Notes in Economics and Mathematical Systems*, Vol. 645, pp. 253–264. Heidelberg, Berlin: Springer-Verlag.

Castelfranchi, C., & Falcone, R. (1998). Principles of trust for MAS: Cognitive anatomy, social importance and quantification. In *Proceedings of ICMAS'98*, pp. 72–79. AAAI Press, Paris.

Castelfranchi, C., Falcone, R., & De Rosis, F. (2001). Deceiving in GOLEM. In C. Castelfranchi & Y. H. Tan (Eds.), *Trust and deception in virtual societies*. Dordrecht, The Netherlands: Kluwer.

de Rosis, F., Pelachaud, C., & Poggi, I. (2004). Transcultural believability in embodied agents. In S. Payr & R. Trappl (Eds.), *Agent culture*. Mahwah, NJ: Lawrence Erlbaum Associates.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Gilbert, N. (2008). *Agent-based models*. Los Angeles: Sage.

Guo, R. (2004). How culture influences foreign trade: Evidence from the U.S. and China. *Journal of Socio-Economics, 33*, 785–812.

Hall, E. T. (1976). *Beyond culture*. Garden City, NY: Anchor.

Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G. J. (2006). Intrinsic and enforceable trust: A research agenda. In *Proceedings of the 99th European Seminar of the EAAE*. Bonn, Germany: Universität Bonn-ILB Press.

Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: Software of the mind* (2nd ed.). New York: McGraw-Hill.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). New York: McGraw-Hill.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2008a). Modeling culture in trade: Uncertainty avoidance. In *2008 Agent-Directed Simulation Symposium (ADSS'08)*. San Diego, CA: SCS.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2008b). Individualism and collectivism in trade agents. In N. T. Nguyen et al. (Eds.), *IEA/AIE 2008* (LNAI 5027), pp. 492–501. Heidelberg, Germany: Springer.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2008c). Long-term orientation in trade. In K. Schredelseker & F. Hauser (Eds.), *Complexity and artificial markets*, Lecture Notes in Economics and Mathematical Systems 614, pp. 107–118. Heidelberg, Germany: Springer.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2009a). Modeling power distance in trade. In N. David & J. S. Sichman (Eds.), *MABS 2008* (LNAI 5269), pp. 1–16. Heidelberg, Germany: Springer.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2009b). Simulation of effects of culture on trade partner selection. In C. Hernández et al. (Eds.), *Artificial economics*. Lecture Notes in Economics and Mathematical Systems 631, pp. 257–268. Heidelberg, Germany: Springer.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2010a). Cultural differentiation of negotiating agents. *Group Decision and Negotiation*. doi:10.1007/s10726-010-9190-x.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2010b). A cross-cultural multi-agent model of opportunism in trade. In N. T. Nguyen & R. Kowalczyk (Eds.), *Transactions on computational collective intelligence II* (LNCS 6450), pp. 24–45. Heidelberg, Germany: Springer.

Hofstede, G. J., Jonker, C. M., & Verwaart, T. (2011). Computational modeling of culture's consequences. In T. Bosse et al. (Eds.), *MABS 2010* (LNAI 6532), pp. 136–151. Heidelberg, Germany: Springer.

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*, 52–88.

Hofstede, G. J., & Pedersen, P. B. (1999). Synthetic cultures: Intercultural learning through simulation gaming. *Simulation & Gaming, 30*, 415–440.

Hofstede, G. J., Pedersen, P. B., & Hofstede, G. (2002). *Exploring culture*. Yarmouth, ME: Intercultural Press.

Hofstede, G. J., Spaans, L., Schepers, H., Trienekens, J., & Beulens, A. (2004). *Hide or confide: The dilemma of transparency*. 's-Gravenhage, The Netherlands: Reed Business Information.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Leadership culture and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

Hwang, P., & Burgers, W. P. (1999). Apprehension and temptation: The forces against cooperation. *Journal of Conflict Resolution, 43*, 117–130.

Jager, W., & Mosler, H. J. (2007). Simulating human behavior for understanding and managing environmental dilemmas. *Journal of Social Issues, 63*(1), 97–116.

Jennings, N. R., Sierra, C., Sonenberg, L., & Tambe, M. (Eds.) (2004). *Proceedings of the third international joint conference on autonomous agents and multi-agent systems, AAMAS'04* (pp. 672–679). New York: IEEE Computer Society Press.

Jonker, C. M., & Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In F. J. Garijo & M. Boman (Eds.), *MAAMAW'99* (LNCS 1647), pp. 221–232. Heidelberg, Germany: Springer.

Jonker, C. M., & Treur, J. (2001). An agent architecture for multi-attribute negotiation. In B. Nebel (Ed.), *Proceedings of the 17th international joint conference on AI, IJCAI'01*, pp. 1195–1201. Seattle, Washington, Morgan Kaufman.

Kashima, Y. (2009). Culture comparison and culture priming: A critical analysis. In R. S. Wyer, C.-Y. Chiu, & Y.-Y. Hong (Eds.), *Understanding culture*. New York: Psychology Press.

Kersten, G. E. (2002). Do E-business systems have culture and should they have one? In *Proceedings of the 10th European Conference on Information Systems, ECIS 2002*, pp. 1237–1248. Gdansk, Poland.

Kersten, G. E., Köszegi, S. T., & Vetschera, R. (2002). The effects of culture in anonymous negotiations: Experiment in four countries. In *Proceedings of the 35th HICSS,* pp. 27. Big Island, HI.

Kirkman, B. R., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37*, 285–320.

Klein Woolthuis, R., Hillebrand, B., & Nooteboom, B. (2005). Trust, contract and relationship development. *Organization Studies, 26*, 813–840.

Kónya, I. (2006). Modeling cultural barriers in international trade. *Review of International Economics, 14*(3), 494–507.

Lévi-Strauss, C. (1992). *Tristes Tropiques*. New York: Penguin Books. (Originally in French, 1955)

Meijer, S. (2009). *The organisation of transactions: Studying supply networked using gaming simulation*. Dissertation, Wageningen University, Wageningen.

Meijer, S., & Verwaart, T. (2005). Feasibility of multi-agent simulation for the trust-and-tracing game. In M. Ali & F. Esposito (Eds.), *IEA/AIE 2005* (LNAI 3533), pp. 145–154. Heidelberg, Germany: Springer.

Minkov, M. (2007). *What makes us different and similar: A new interpretation of the world values survey and other cross-cultural data*. Sofia, Bulgaria: Klasika I Stil.

Modelling trade and trust across cultures. In Stølen, K., Winsborough, W. H., Martinelli, F., Massacci, F. (Eds.) (2006, May 16–19). *Proceedings of the 4th International Conference on Trust Management*, iTrust 2006. Pisa, Italy. *Proceedings, Series: Lecture Notes in Computer Science* (Vol. 3986, pp. 120–134).

Olekalns, M., & Smith, P. L. (2009). Mutually dependent: Trust, affect and the use of deception in negotiation. *Journal of Business Ethics, 85*, 347–365.

Payr, S., & Trappl, R. (2004). *Agent culture; human-agent interaction in a multicultural world*. Mahwah, NJ: Lawrence Erlbaum Associates.

Raser, J. C. (1969). *Simulations and society: An exploration of scientific gaming*. Boston, MA: Allyn & Bacon.

Rehm, M., André, E., Bee, N., Endrass, B., Wissner, M., Nakano, Y. I., et al. (2007). The CUBE-G approach – Coaching culture-specific nonverbal behavior by virtual agents. In *Proceedings of the 38th conference of the international simulation and gaming association*, Nijmegen, The Netherlands.

Rehm, M., Nishida, T., André, E., & Nakano, Y. I. (2008). Culture-specific first meeting encounters between virtual agents. In *IVA 2008* (LNCS 5208). Heidelberg, Germany: Springer.

Schimmack, U., Oishi, S., & Diener, E. (2005). Individualism: A valid and important dimension of cultural differences between nations. *Personality and Social Psychology Review, 9*, 17–31.

Schwarz, S. H. (2009). Culture matters: National value cultures, sources, and consequences. In R. S. Wyer et al. (Eds.), *Understanding culture*. New York: Psychology Press.

Silverman, B. G., Bharathy, G., Johns, M., Eidelson, R. J., Smith, T. E., & Nye, B. (2007). Sociocultural games for training and analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part A, 37*, 1113–1130.

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: Part I: Enabling science with PMFserv. *Presence, 15*, 139–162.

Smith, P. B. (2002). Culture's consequences: Something old and something new. *Human Relations, 55*, 119–135.

Smith, P. B. (2006). When elephants fight, the grass gets trampled: The GLOBE and Hofstede projects. *Journal of International Business Studies, 37*, 915–921.

Smith, P. B., Dugan, F., & Trompenaars, F. (1996). National culture and the values of organizational employees: A dimensional analysis across 43 nations. *Journal of Cross-Cultural Psychology, 27*, 231–246.

Sóndergaard, M. (1994). Hofstede's consequences: A study of reviews, citations and replications. *Organization Studies, 15*, 447–456.

Steinel, W., & De Dreu, K. W. (2004). Social motives and strategic misrepresentation in social decision making. *Journal of Personality and Social Psychology, 86*, 419–434.

Tönnies, F. (1963). *Community and society*. New York: Harper & Row. (Original pub. 1887)

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, MA: Westview.

Triandis, H.C., Carnevale, P., Gelfand, M., Robert, C., Arzu Wasti, S., Probst, T. (2001). Culture and deception in business negotiations: A multilevel analysis. *International Journal of Cross Cultural Management, 1*, 73–90.

Trompenaars, F. (1993). *Riding the waves of culture*. London: Nicholas Brealy.

Tykhonov, D., Jonker, C., Meijer, S., & Verwaart, T. (2008). Agent-based simulation of the trust and tracing game for supply chains and networks. *Journal of Artificial Societies and Social Simulation, 11*(3), 1.

Ward, D., & Hexmoor, H. (2003). Towards deception in agents. In *Proceedings of AAMAS 2003*, pp. 1154–1155. Melbourne, Australia.

Weisbuch, G., Kirman, A., & Herreiner, D. (2000). Market organisation and trading relationships. *Economic Journal, 110*, 411–436.

Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: Free Press.

Williamson, O. E. (1998). Transaction cost economics: How it works; where it is headed. *De Economist, 146*, 23–58.

Wilson, D. S. (2007). *Evolution for everyone*. New York: Delacorte Press.

Wirtz, J., & Kum, D. (2004). Consumer cheating on service guarantees. *Journal of the Academy of Marketing Science, 32*(2), 159–175.

How Do I Trust Thee? Dynamic Trust Patterns and Their Individual and Social Contextual Determinants

C. Ashley Fulmer and Michele J. Gelfand

Abstract Levels of trust in a social relationship can fluctuate over time. However, few studies have directly examined the different trust decisions involved as the relationship progresses and delineated the various trust patterns that arise from these decisions. Our analysis examined the key processes in trust decisions during the phases of trust formation, dissolution, and restoration. These trust decisions across phases give rise to trust profiles. In this chapter, we discuss six unique trust profiles: High Trust, High Distrust, Tit-for-Tat, Seizing and Freezing, Assessment, and Grim Trigger, based on divergent trust patterns across trust phases. We focus on how individual and social contextual factors affect these patterns of trust. Our model highlights the need to consider multiple trust phases both theoretically and empirically to offer a more holistic understanding of how trust unfolds overtime in social relationships.

Keywords Trust • Trust building • Trust violation • Time • Profiles • Individual characteristics • Social context • Culture • Dynamic analysis

C.A. Fulmer (✉)
Department of Psychology, National University of Singapore, Singapore
e-mail: afulmer@nus.edu.sg

M.J. Gelfand
Department of Psychology, University of Maryland, College Park, MD, USA

1 Introduction

Consider the following examples:

> An employee agrees to be in a team project and puts in considerable amounts of time and effort. However, when she hears that her colleagues might have taken the credit for her work, she immediately decides to withdraw from the team project and requests a transfer to a different branch to avoid working with this group of colleagues in the future.
>
> A college graduate starts a business with an acquaintance whom he met only recently. Despite finding out that the person repeatedly mismanages their accounts, he continues to entrust the person to handle this vital function of the business, believing that the incidents of mismanagement are accidental and beyond the person's control because of a difficult economic climate.
>
> A buyer negotiates a partnership with a seller in a different country. Although during the negotiation the buyer finds the seller to be forthright and fair, he remains distrusting of him and watchful for potential contract violations. The partnership nearly breaks down when a shipment from the seller is delayed. The buyer believes that the delay is a sign of incompetence, but reluctantly agrees to give the seller another chance after receiving multiple apologies from the seller. The partnership continues but remains fragile.

As illustrated in these examples, trust can fluctuate widely over the course of a relationship as people manage their interdependence. A large body of research has indicated that trust is critical in all stages of relationships, from their initial development to breakdown and reestablishment. Research across a wide range of disciplines, including psychology, sociology, and economics, has found that trust serves vital functions. It promotes cooperation and teamwork (Dirks, 1999; Jarvenpaa & Leidner, 1999; Simons & Peterson, 2000), organizational citizenship behavior (Connell, Ferres, & Travaglione, 2003; McAllister, 1995; Podsakoff, MacKenzie, Moorman, & Fetter, 1990), leadership effectiveness (Dirks & Ferrin, 2002; Gillespie & Mann, 2004; Jung & Avolio, 2000), success in negotiation and mediation (Arnold & O'Connor, 2006; Olekalns, Lau, & Smith, 2002; Valley, Moag, & Bazerman, 1998), and mergers and acquisitions (Graebner, 2009; Maguire & Phillips, 2008; Stahl & Sitkin, 2005), as well as national-level democracy (Putnam, 1993) and economic well-being (Fukuyama, 1995).

The examples above highlight the fact that trust in a social relationship evolves and changes over time. A trust relationship can go through different phases, including *trust formation*—a progression in which individuals choose to trust others and increase their trust over time; *trust dissolution*—a progression which occurs after violations when individuals decide to lower their trust in others; and *trust restoration*—a progression which occurs when trust stops declining after violation and rebounds toward stability (e.g., Kim, Dirks, & Cooper, 2009; Miles & Creed, 1995; Rousseau, Sitkin, Burt, & Camerer, 1998). In each of these trust phases, individuals need to make a trust decision about whether to increase or decrease trust in the trustee. For example, some people can be more reluctant to build trust than others in trust formation. Likewise, during trust dissolution, some individuals may consider the smallest sign a trust breach, while others may give more latitude and the benefit of the doubt before deciding to lower their trust. Finally, people also diverge in their rate of trust recovery. Once people lower their trust in another, some may find it difficult to trust that person again while others are more willing to give the person another chance.

The examples in the beginning of this chapter illustrate highly different trust dynamics reflecting different decisions across phases. The first example of the team project shows a relatively fast trust formation, followed by a rapid trust dissolution and a failure in trust restoration. The second example of friends has an immediate trust formation and very little and slow trust decline despite the presence of multiple trust violations. The third example of a manufacturer-supplier partnership demonstrates a difficult and slow trust formation, a fast dissolution, and a slow trust restoration.

Despite the recognition that multiple trust phases exist in relationships, research has overwhelmingly focused on a single trust phase at a time (Lewicki & Bunker, 1996; Rousseau et al., 1998), and empirical and theoretical work on the trajectories of trust across phases remains limited. Yet an examination of trust phases in isolation cannot delineate how trust patterns unfold as people interact. An understanding of changes in trust patterns over time is particularly important as scholars have agreed that trust is dynamic (Rousseau et al.). By considering multiple phases sequentially and studying the resultant trust patterns, we can predict the individual and social contextual factors that might produce the different trust *dynamics*.

The goals of this chapter, therefore, are to develop a theory regarding trust trajectories—different dynamics that occur across trust formation, dissolution, and restoration—and to identify the individual and social-contextual predictors of different trust trajectories. We first discuss each of the three trust phases in the next section, followed by an examination of a series of dynamic trust patterns produced by a simultaneous consideration of these three phases. Examples of individual and social contextual determinants of each trust profile are highlighted. We end the chapter with empirical and practical considerations that can be derived from our theoretical analysis.

2   Phases of Trust

The trust literature has identified three distinct phases of trust—formation, dissolution, and restoration (Kim et al., 2009; Miles & Creed, 1995; Rousseau et al., 1998). These trust phases are interrelated and yet each has its distinct properties, processes, and functions. Reviewing the interdisciplinary research on trust, including psychology, sociology, and economics, Rousseau and colleagues concluded that trust is neither static nor stable. Instead, researchers should examine trust dynamically across phases as it fluctuates and changes in response to individual factors and social contextual demands. Before our analysis of the profiles of different trust patterns, each of the three trust phases is discussed in turn below.

2.1   Trust Formation

A trust relationship begins with trust formation (Cook et al., 2005; Mayer, Davis, & Schoorman, 1995; McKnight, Cummings, & Chervany, 1998). The key process

in trust formation is how trustors can infer the trustworthiness of the trustee. The perception of trustworthiness is multi-faceted and can be categorized in the forms of the ability, benevolence, and integrity of the trustee (Mayer et al., 1995). Ability refers to the trustee's competence to carry out an expected action, benevolence refers to the trustee's intrinsic and positive intention toward the trustor, and integrity refers to the trustor's perception that "the trustee adheres to a set of principles that the trustor finds acceptable" (Mayer et al., p. 719). Integrity also includes factors such as honesty, fairness, consistency, and reliability (Mayer et al.).

To the extent that trustors perceive the trustee is sufficiently competent, benign, and upright, they are more likely to consider the trustee to be trustworthy. As a result, they are more likely to decide to increase their trust quickly during trust formation. Note that depending on the type of relationship and who the trustee is, trust formation can occur without all three of the trustworthiness facets. For example, in a buyer-supplier relationship, a positive perception of the other party's ability and integrity may be sufficient to build trust that sustains the partnership without an explicit consideration of benevolence. As the relationship progresses and the number of interaction domains between the two parties increases, however, all three facets of trustworthiness are likely to be required due to the growing breadth of the relationship (Lewicki, McAllister, & Bies, 1998). Researchers have observed incidents in which organizations in stable alliances will forgo better business opportunities with alternative partnerships in exchange for the benevolence, commitment, and loyalty of their long-term partners (Yamagishi & Yamagishi, 1994).

2.2   Trust Dissolution

The very conditions that foster trust allow for future trust breaches (Granovetter, 1985). To trust, in essence, is to take risk based on positive expectations of others. This puts the trustor in a position vulnerable to violations. More frequently than we expect, violations occur in a trusting relationship (Elangovan & Shapiro, 1998; Kim et al., 2009; Morris & Moberg, 1994; Robinson & Bennett, 1995). The second phase of trust—dissolution—concerns decreases in trust after one or more violations have occurred. Two key processes in this phase, vigilance and idiosyncratic credits, affect people's decisions on whether to lower trust. Borrowing from the definitions of the signal detection theory in the field of human factor and applied cognition (Blough, 2001; Parasuraman, 1998; Wickens & Hollands, 2000) and sensitivity to rejection in the close relationship literature (Downey & Feldman, 1996), we define vigilance as sustained attention and alertness directed to detect and identify signals of intentional violation of trust. When a trustor is vigilant toward violations, he or she will be more likely to notice violations and lower trust readily.

Once a violation has been identified, trustors need to make a decision about how much trust to lower in the trustee. However, an awareness of violation or injustice may not always prompt responding actions (Greenberg & Alge, 1998). Some may be likely to reduce trust immediately, whereas others might be willing to withhold

such decisions if they grant the other party "idiosyncratic credits." Idiosyncratic credit is defined as an accumulation of positively disposed impressions residing in the perceptions of relevant others (Hollander, 1958, p. 120). Idiosyncratic credits are similar to the concept of latitude of acceptance in social judgment theory, defined as the range of positions accepted and tolerated (Hovland, Harvey, Sherif, 1957; Sherif & Hovland, 1961).

Both idiosyncratic credits and latitude of acceptance suggest that individuals may not change their trust attitude and lower their trust based on a single violation, but rather a number of violations. Research on justice violation threshold, for example, shows that individuals vary in the number of violations they allow before responding to injustice (Beugré, 2005; Gilliland, Benson, Schepers, 1998). To the extent that a trustor allows high levels of idiosyncratic credits for the trustee, the trustor should be less likely to lower trust at the first sight of a potential violation. The trustee is placed under "pardons and paroles" instead so that the social relationship can continue to function (Doz & Hamel, 1998). The rate of trust dissolution should thus depend on the vigilance level of a trustor and the idiosyncratic credits the trustor gives to a trustee.

2.3   Trust Restoration

The third phase of trust is restoration. Although violations can be a part of a trusting relationship, there is variation in how much and how fast the betrayed individuals restore their trust in the other party. A key process that affects trust restoration is the trustor's attribution of the violation that has occurred. In the forgiveness literature, intent, controllability, and stability have been proposed as three major aspects within one's causal attribution of a negative event in a relationship (Fincham, 2000; Heider, 1958; Weiner, 1986, 1995). Specifically, forgiveness and trust restoration have been predicted to be more likely when the trustor believes the cause of a violation to be unintentional and due to the situation, uncontrollable on the part of the trustee, or an isolated incident and unstable characteristic of the trustee (Tomlinson & Mayer, 2009). A recent meta-analysis by Fehr, Gelfand, and Nag (2010) indeed showed the negative effects of intent (weighted mean $r = -.49$) and responsibility (weighted mean $r = -.35$) on forgiveness.

Individual differences can also affect attributions. For example, some individuals can be high on hostile attributional bias (Dodge, Price, Bachorowski, & Newman, 1990), a tendency to make negative attributions of others' intentions in ambiguous situations. These trustors would tend to think that a violation is intentional, controllable, and an action typical of the trustee. In contrast to this negative preconception, prior positive relationship history and empathy can buffer the trustors and prompt them to make positive attributions of the violations (Fehr et al., 2010). These examples underlie the fact that the attributional process during trust restoration is subject to selective interpretation (Fiske & Taylor, 1984).

3   Defining Trust Trajectories

As the above discussion implies, when a trustor makes the trust decision across the three trust phases, levels of trust can fluctuate at divergent rates in different phases. In each trust phase—formation, dissolution, and restoration—levels of trust can change quickly or slowly. The trust changes across phases form a dynamic trust pattern, or what we refer to as *trust trajectories*.

More specifically, a trust trajectory indicates the magnitude of changes in trust level from one point in time to another. A trajectory is the same as the slope in a regression. On a graph, trajectory is shown by the differences in elevation among phases of trust. A small difference indicates a gentle slope and *slow* trust changes, whereas a large difference indicates a steep slope and *fast* trust changes. For example, as Fig. 1 displays, a fast trust formation refers to a large increase in trust levels during the formation phase. This is followed by a slow trust dissolution with a small decrease in trust levels in the dissolution phase, and a fast trust restoration leading to a quick rise in trust levels over time.

When we consider multiple trust phases sequentially, different trust patterns arise. We produced the different profiles of trust patterns by roughly differentiating trust changes in each of the three phases as either fast or slow and crossing the three trust phases orthogonally ($2 \times 2 \times 2$). This results in eight possible trust profiles that have yet to be identified in the literature. These trust profiles, each with unique trajectories across the formation, dissolution, and restoration phases, form the basis of our theoretical propositions.
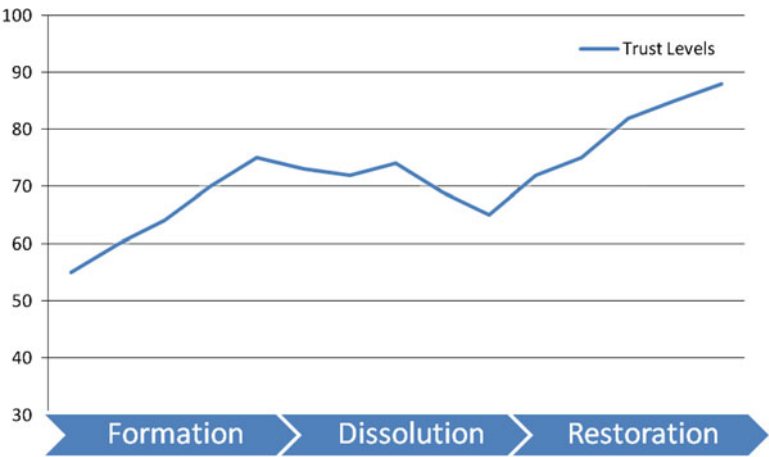
Fig. 1 The High Trust profile with a fast formation, a slow dissolution, and a fast restoration

In this chapter, we focused primarily on six profiles.[1] Specifically, *high trust profiles* are characterized by fast formation, slow dissolution, and fast restoration. In contrast, *high distrust profiles* are characterized by slow formation, fast dissolution, and slow restoration. *Tit-for-tat trust profiles* are characterized by fast formation, fast dissolution, and fast restoration, and *seizing and freezing trust profiles* are characterized by fast formation, slow dissolution, and slow restoration. *Assessment trust profiles* are characterized by slow formation, slow dissolution, and slow restoration. Finally, *grim trigger trust profiles* are characterized by fast formation, fast dissolution, and slow restoration.

In what follows, we discuss each of these profiles in more detail and we delineate examples of specific individual and social contextual differences that are theorized to underlie these different dynamics. Table 1 provides an overview of our discussion. We note that the list of variables identified for each profile is by no means exhaustive. They are examples to illustrate how various factors can affect the key process in each trust decision and the trust patterns across phases.

3.1 High Trust Profile: Fast Formation, Slow Dissolution, Fast Restoration

As can be seen in Fig. 1, this trust profile is characterized by a rapid trust formation in the beginning of a trust relationship, a slow trust dissolution when violations have occurred, and a quick trust restoration after dissolution. Fundamental to this profile is the notion that trustors do not perceive the trusting relationships to be risky and that their trust is very resilient, thus enabling them to 'move on' after trust violations. The individual differences relevant to this profile include high self-esteem (Leary, Tambor, Terdal, & Downs, 1995), secure attachment style (Mikulincer, 1998), and locomotion regulatory mode (Kruglanski et al., 2000). The social contextual factors important to this profile are relational embeddedness (Fehr et al., 2010; Lewicki & Bunker, 1995, 1996), common third party and reputation (Ferrin, Dirks, & Shah, 2006; Glick & Croson, 2001), and collectivism (with ingroups; Triandis, 1995). Collectively these variables tend to facilitate perception of trustworthiness of others, lower vigilance toward violations, increase idiosyncratic credits that the trustors are willing to give to the trustees, and promote positive attributions of the violations.

First, a high level of self-esteem should lead to the high trust pattern. Rotter (1954, 1967, 1971) discussed trust from a social learning perspective, explaining that trust is a result of one's generalized expectancies of the trustworthiness of other people that one developed over time. As individuals high on self-esteem tend to have

---

[1]There are two additional possible trust profiles: (1) slow formation, fast dissolution, and fast restoration and (2) slow formation, slow dissolution, fast restoration. However, because not all trust profiles occur with equal frequency, we focused on the six commonly occurred trust profiles in this chapter.

Table 1 The trust patterns and the individual and social contextual determinants

Trust profiles	Individual and social contextual determinants	
Profile #1 High Trust		
Fast formation; slow dissolution; fast restoration	Individual	High self-esteem
		Secure attachment
		Locomotion
	Social contextual	Relationship history
		Third party and reputation
		Collectivism with ingroup
Profile #2 High Distrust		
Slow formation; fast dissolution; slow restoration	Individual	Low self-esteem
		Anxious attachment
		Cynicism
		Betrayal aversion
	Social contextual	Surveillance and monitoring
		Honor logics
		Collectivism with outgroup
Profile #3 Tit-for-Tat Trust		
Fast formation; fast dissolution; fast restoration	Individual	Tit-for-tat
		Equity sensitivity
	Social contextual	Exchange relationships
		Dignity logics
Profile #4 Seizing and Freezing Trust		
Fast formation; slow dissolution; slow restoration	Individual	Need for closure
	Social contextual	Time Pressure
		Power distance with authority
Profile #5 Assessment Trust		
Slow formation; slow dissolution; slow restoration	Individual	Assessment
		Need to avoid closure
	Social contextual	Public situation
Profile #6 Grim Trigger Trust		
Fast formation; fast dissolution; slow restoration	Individual	Grim strategy
		Interpersonal orientation
	Social contextual	Performance orientation

a positive expectation of interactions with others (Leary et al., 1995), the process of trust formation for them should be fast. Further, verification of the self by another person, which occurs more frequently with individuals with high (vs. low) self-esteem, can also facilitate trust formation (Burke & Stets, 1999). High self-esteem individuals have been found to be less concerned about threats to their relationship (Murray, Holmes, & Collins, 2006) and they should allow more idiosyncratic credits in trust dissolution. During trust restoration, their positive expectations of others (Leary et al., 1995) should prompt them to make positive attributions of the violations, believing that the violations are unintentional, uncontrollable, and uncharacteristic of the trustee.

Related to self-esteem is the construct of secure attachment (Cassidy, 1988). The secure attachment style in adults has consistently been linked to a higher level of trust in others (Collins & Read, 1990; Hazan & Shaver, 1987; Mikulincer, 1998). Specifically, people with a secure attachment style should have greater perceptions of the benevolence of trustees, and expect that the trustees care about their personal well-being (McAllister, 1995). Murray and colleagues (2006) also found that adults with a secure attachment style were less concerned about social rejection and violations. Likewise, Mikulincer (1998) found that adults with a secure attachment style felt more trust in partners in close relationships and adopted constructive coping strategies such as communication after violations. Their positive perceptions of the relationship and partner should also promote positive attributions of the violations (Collins, Ford, Guichard, & Allard, 2006). For example, Collins and colleagues (2006) found that secure trustors were less likely to make intentionality attributions than anxious trustors. Based on these factors, we predict that secure attachment should lead to fast formation, slow dissolution, and fast restoration.

Motivational factors are also relevant to this trust pattern. In particular, individuals who endorse self-regulation strategies of locomotion (Kruglanski et al., 2000) can exhibit the high trust profile. Locomotors are "concerned with movement from state to state and with committing the psychological resources that will initiate and maintain goal-related movement in a straightforward and direct manner, without undue distractions or delays" (Kruglanski et al., 2000, p. 794). It has been noted that interpersonal relationships can serve as a means to a goal through which motivation and goal commitment can be elicited (Kruglanski et al., 2002). Consequently, when locomotors enter a trusting relationship instrumental to a goal, they should build trust quickly. Another unique quality of locomotors is their willingness to invest efforts (Pierro, Kruglanski, & Higgins, 2006). They continually exhibit goal-directed persistency despite setbacks (Shah & Kruglanski, 2003). When facing violations, therefore, locomotors are expected to lose trust slowly and repair trust quickly to maintain relationships in order to continue the tasks at hand. Research has also found that locomotors in close relationships frequently engage in affirmation of the partner (Kumashiro, Rusbult, Finkenauer, & Stocker, 2007). Based on these considerations, we expect that when pursuing a goal, high locomotors will perceive the trustee to be more trustworthy, behave in a less vigilant manner, and make more positive attributions of violations than low locomotors.

At the social contextual level, the degree to which individuals are embedded in relationships that are of high quality—with satisfaction, commitment, trust, and communication (Canevello & Crocker, 2010; Collins & Read, 1990)—is related to this trust profile. A pre-existing positive relationship should facilitate trust building (Jones & George, 1998; Lewicki & Bunker, 1995, 1996). During the trust dissolution phase, previous positive relational experiences with another can decrease vigilance and reaction to violations (Murray et al., 2006). A trustor's perceived idiosyncratic credit for a trustee can similarly be affected by the history of the particular relationship. For example, if the trustor is in a committed relationship of high quality, the trustor should be more likely to grant more idiosyncratic credits

to the trustee. Positive relationship quality should also promote trust restoration. In a meta-analysis on forgiveness, Fehr and colleagues (2010) have found relationship closeness, satisfaction, and commitment between a trustor and trustee to be strong determinants of forgiveness. These signs of high relationship quality should motivate a trustor to overcome negativity to maintain a trust relationship (Finkel, Rusbult, Kumashiro, & Hannon, 2002; McCullough et al., 1998).

In cases where the trustor and trustee do not know each other directly, a trusted third party between the trustor and trustee can lead to a high trust pattern. During trust formation, a common third party may help the trustor assess the trustee's trustworthiness. Ferrin and colleagues (2006) found that a trusted third party directly increases the trustor's trust in the trustee. We expect that the effect of a common third party should be even stronger if the third party, an ingroup member of the trustor, lends the perception that the trustee is also an ingroup member of the trustor. People tend to perceive a higher level of trustworthiness from an ingroup member compared to an outgroup member (Brewer & Kramer, 1985; Kramer & Brewer, 1984). Sharing a common identity with the trustee, such as one through common group membership, has been found to facilitate trustor's trust building (Buchan, Croson, & Dawes, 2002). In this way, a common third party should also benefit the process of trust dissolution and restoration by lowering vigilance, increasing idiosyncratic credits, and leading to a positive attribution of a violation.

When a common third party is not available, the trustor may nevertheless rely on the reputation of the trustee (Glick & Croson, 2001) to infer the ability, benevolence, and integrity of the trustee. Reputation has been found to increase trust, particularly in a competitive environment (Glückler & Armbrüster, 2003). A positive reputation should also lead the trustor to have lower vigilance and higher idiosyncratic credits during trust dissolution, as well as more positive attributions of the violation during trust restoration.

Culture is another relevant social contextual variable to this pattern. In particular, when people from collectivistic cultures interact with ingroup members, they should exhibit the high trust pattern. Collectivism is defined as "a situation in which people belong to in-groups or collectivities which are supposed to look after them in exchange for loyalty" (Hofstede, 1984, p. 419). Therefore, the distinction of ingroup versus outgroup status of others is especially relevant to collectivists (Triandis, 1995; Triandis, McCusker, & Hui, 1990). As mentioned previously people tend to perceive their ingroup members to be trustworthy (Brewer & Kramer, 1985; Kramer & Brewer, 1984). This effect should be stronger among collectivists and lead to a fast trust formation. We similarly predict that the higher cohesiveness with ingroup members (Hofstede) will lead collectivists to allow more idiosyncratic credits for their ingroup trustee (particularly when violations are not large) and exhibit a slower trust dissolution. Finally, because collectivists are more likely to make situational attributions (Miller, 1984; Morris & Peng, 1994) and have higher perspective taking (Cohen & Hoshino-Browne, 2005; Kitayama & Markus, 1999), they are likely to make positive attributions of the violations and display a fast trust restoration.
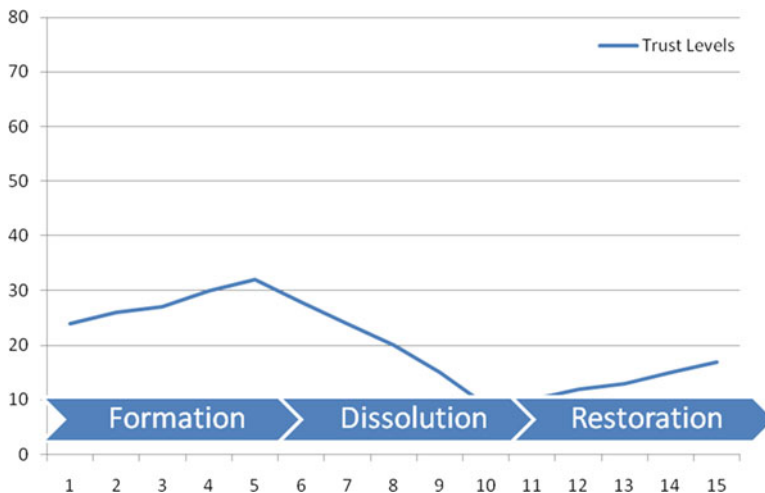
Fig. 2 The high distrust profile with a slow formation, a fast dissolution, and a slow restoration

3.2 High Distrust Profile: Slow Formation, Fast Dissolution, and Slow Restoration

As shown in Fig. 2, a high distrust profile is characterized by a prolonged trust formation in the beginning of a relationship, a rapid trust dissolution when violations have occurred, and a slow and difficult trust restoration after dissolution. Fundamental to this profile is the notion that trustors perceive the decisions to trust and social relationships to be very risky and that their trust falters easily. The individual differences relevant to this profile include low self-esteem (Leary et al., 1995), anxious attachment style (Mikulincer, 1998), cynicism (Costa, Zonderman, McCrae, & Williams, 1985), and betrayal aversion (Bohnet & Zeckhauser, 2004). The social contextual factors important to this profile include surveillance and monitoring (Sitkin & Roth, 1993) and, at a more macro level, cultures of honor (Nisbett & Cohen, 1996) and collectivism (with outgroups; Triandis, 1995). Overall, these variables decrease perception of trustworthiness of others, heighten vigilance toward violations, limit idiosyncratic credits granted to the trustees, and lead to negative attributions of the violations.

Opposite to the high trust profile, individuals with low self-esteem and anxious attachment style should exhibit a high distrust profile. People with low self-esteem tend to have more negative expectations of others, resulting in a low perception of trustworthiness and a slow trust formation. The chronic negative expectations of other people should also lead them to expect rejections and violations in relationships (Leary et al., 1995), be highly vigilant (Murray et al., 2006), and allow few idiosyncratic credits. As a result, trust should decrease quickly at the first likely sign of violation. Finally, the negative expectations people with low self-esteem

have about others should lead them to make negative attributions of the violations, believing they are intentional, controllable, and characteristic of the trustee.

Compared to individuals with a secure attachment style, those with an anxious attachment style tend to have a lower level of trust (Collins & Read, 1990; Hazan & Shaver, 1987; Mikulincer, 1998). It stands to reason that trust formation can be difficult for them. The anxious attachment style has also been found to lead to hypervigilance toward threatening cues in a relationship (Murray et al., 2006; Shaver & Mikulincer, 2003). We also expect that individuals with an anxious attachment style would allow few idiosyncratic credits and exhibit fast trust dissolution. After violations, anxious individuals tend to engage in ruminative worry (Mikulincer). Rumination, defined as "a passive and repetitive focus on the negative and damaging features of a stressful transaction" (Skinner, Edge, Altman, & Sherwood, 2003, p. 242), has been found to lower forgiveness (Fehr et al., 2010). Anxious individuals have also been found to endorse more relationship-threatening and conflict-inducing attributions than secure individuals (Collins et al., 2006). Consequently, trust restoration can be slow for people with low self-esteem.

In addition to self-esteem and attachment styles, another individual-difference variable that can affect people's trust decisions throughout trusting relationships is their levels of cynicism. Cynicism reflects "distrusting and disparaging attitudes toward the motives of others, and beliefs in the selfishness of human nature" (Costa et al., 1985, p. 929; see also Leung et al., 2002). Notably, cynicism has been differentiated from distrust on multiple aspects, such as the emotions of disgust and shame that are unique to the former and the requirement of vulnerability to the latter (Dean, Brandes, & Dharwadkar, 1998). Nevertheless, we expect that the negative view of others prescribed by cynicism will lead individuals to display the high distrust pattern. In organizations, for example, cynics have been found to be more likely to question the motives of their leaders (Kanter & Mirvis, 1989). As a result, people high on cynicism should have lower perceptions of others' ability, integrity, and benevolence. Cynics are also expected to have high vigilance toward violations and grant few idiosyncratic credits, especially as cynical attitudes are typically acquired though a series of unmet expectations and disappointment (Andersson, 1996) and a lack of contingency between one's actions and the environment (Triandis, 1976). Because cynicism is related to a host of negative emotions such as distress and disgust (Andersson & Bateman, 1997; Dean et al., 1998) and a belief that others are selfish (Costa et al., 1985), individuals high on cynicism should be more likely to make negative attributions after violations than those low on cynicism, making trust restoration slower and more difficult.

Individuals also differ in the extent to which they are willing to take risk from other people. Bohnet and her colleagues (2004, 2008) termed the trustors' tendency to avoid violations from other individuals as betrayal aversion. They suggested that it is betrayal aversion, not risk aversion alone, that affects people's trust decisions (Bohnet & Zeckhauser, 2004). More specifically, betrayal aversion refers to a phenomenon wherein individuals are more averse toward social risk caused by other people than they are toward natural risks from inanimate objects (Bohnet, Greig, Herrmann, & Zeckhauser, 2008). This is partially because people

are concerned with process through which their outcomes are generated, which is a distinct concern when dealing with human versus inanimate decision-making (Rabin, 1993).

When individuals are betrayal averse, they should thus be less likely to find another person trustworthy in the beginning of a trust relationship, more likely to be highly vigilance and assume a violation has occurred, and more likely to make negative attributions of the trust violation. Empirical evidence indeed suggested that people tend to be more conservative in their risk-taking when the outcomes are determined by another individual (i.e., betrayal averse) than by randomness (i.e., risk averse), even though the outcomes are identical (Bohnet et al., 2008; Bohnet & Zeckhauser, 2004; Fehr, 2009; Koehler & Gershoff, 2003). Failing to consider the interpersonal component in the examination of trust as risky decisions may explain some of the mixed empirical results on the relationship between trust and risk (*c.f.* Bohnet & Zeckhauser, 2004).

Highly relevant to the individual differences of betrayal aversion is the social contextual factor of honor cultures. Honor has been defined as "the value of a person in his own eyes, but also in the eyes of his society. It is his estimation of his own worth, his claim to pride, but it is also the acknowledgement of that claim" (Pitt-Rivers, 1966, p. 21). Miller (1993) further explained that "honor is above all the keen sensitivity to the experience of humiliation and shame" (p.84). Just as betrayal aversion is high in the Middle East, the culture of honor is pronounced in the region, suggesting that betrayal aversion and the honor logics may be interrelated. Bohnet, Herrmann, and Zeckhauser (2010) found that people in the Persian Gulf regions required a higher level of trustworthiness before they were willing to trust than Swiss and Americans. People in honor cultures may have a stronger concern for the ability, benevolence, and integrity of the trustee than people in other cultures, which can lead to a slow trust formation. Indeed, benevolence has been found to be an especially relevant concern in Middle East cultures (Dorfman, Hanges, & Brodbeck, 2004).

Trust dissolution may likewise be difficult and can occur quickly in honor cultures. Bohnet and colleagues (2008) have found that people in the Middle East have an exceptionally high level of betrayal aversion, compared to Americans, Europeans, and East Asians. As discussed previously, this high level of betrayal aversion should lead to heightened vigilance for potential trust violations. Further, people in honor cultures are compelled to defend their honor and seek revenges when others trespass their properties and rights or fail to reciprocate in a relationship (Nisbett & Cohen, 1996). This is because an individual's personal worth is determined by both the self and others in these cultures (Nisbett & Cohen, 1996; Stewart, 1994). As a result, violations acquire a special meaning in honor cultures as they are a direct test of what others can do to one and an indirect test of one's self worth (Leung & Cohen, 2011). In the justice literature, reactions to injustice lead to stronger reactions when individuals' self-worth is threatened (Greenberg & Alge, 1998). It stands to reason that, following trust violations, trust restoration in honor cultures would be very slow and difficult.

Another relevant social contextual factor to the high distrust profile is collectivism. In particular, people in collectivistic cultures should display the high distrust

pattern when dealing with outgroup members. Collectivists value relationships with ingroup members and perceive large social distance with outgroup members (Hofstede, 1984; Triandis et al., 1990). Further, people tend to find outgroup members to be less trustworthy than ingroup members (Brewer & Kramer, 1985; Kramer & Brewer, 1984), and this effect should be stronger for collectivists. Because of the distrust collectivists have for an outgroup member, it can take longer for them to build trust. Branzei, Vertinsky, and Camp (2007) found that collectivists engage in more incremental testing of a stranger's trustworthiness in a trust relationship than individualists, suggesting that collectivists may be vigilant toward violations and give limited idiosyncratic credits with an outgroup trustee. During trust restoration, we expect that collectivists would make more negative attributions when violations are from an outgroup member than an ingroup member, making the process slower than usual.

Finally, at the social contextual level, surveillance and monitoring can also lead to the high distrust pattern. To ensure trustworthiness, people sometimes elect to monitor their trustees. The degree of monitoring in a society is related to its levels of constraints. People in a tight society, characterized by strong and pervasive social norms and little tolerance toward deviations (Gelfand, Nishii, & Raver, 2006), should engage in and be accustomed to a high level of monitoring than those in a loose society. Tight societies, for example, have been found to have higher population density and higher police, both of which afford higher monitoring (Gelfand et al., 2011).

Monitoring and surveillance can have unintended negative effects on trust. The emancipation theory of trust (Yamagishi & Yamagishi, 1994) suggests that trust is a more effective means to manage social uncertainty than active surveillance and monitoring because trust allows flexibility for parties in a relationship to focus their energy on performing constructive activities instead of safeguarding and checking compliance (c.f. Mayer & Gavin, 2005). Moreover, McAllister (1995) found that trust and monitoring are negatively related, a result consistent with prior literature (Kruglanski, 1970; Strickland, 1958). The efforts that the trustors use in conducting surveillance and monitoring may therefore ironically decrease their efforts to promote trust (Kramer, 1999). Cialdini (1996) suggested additional reasons for this negative relationship, such as attributions of honest behaviors to the surveillance by both the trustor and trustee and increased fear and suspicion in both parties. Fear and suspicion are likely to increase trustors' vigilance and decrease the idiosyncratic credits they are willing to grant. Together, these effects should lead a trustor to perceive a low level of others' trustworthiness, be highly vigilant, and make negative attributions of the violation.

The negative relationship between monitoring and trust could be due to the strong overlap between trust and risk. As trust only occurs when individuals face uncertainty (Gambetta, 1988; Mishra, 1993) and need to make a "leap of faith" choice despite the uncertainty (Holmes & Boon, 1991; Sitkin & Roth, 1993), monitoring eliminates the necessary uncertainty and deprives opportunities for trust to develop. In addition to being detrimental to trust, surveillance and monitoring may actively promote distrust, a confident expectation about the negative attitudes and behaviors of another (Lewicki et al., 1998).
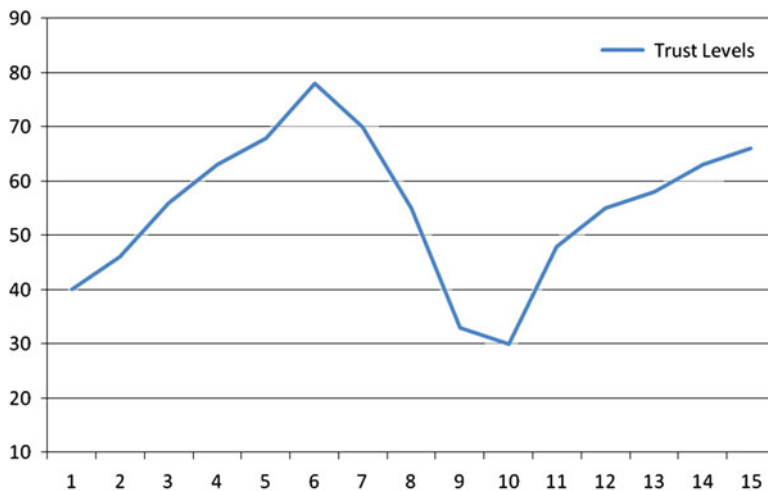
Fig. 3 The tit-for-tat trust profile with a fast formation, a fast dissolution, and a fast restoration

3.3 Tit-for-Tat Trust Profile: Fast Formation, Fast Dissolution, and Fast Restoration

As Fig. 3 demonstrates, the tit-for-tat trust profile is characterized by a quick trust formation in the beginning of a trust relationship, a quick trust dissolution when violations have occurred, and a quick trust restoration after dissolution. Fundamental to this profile is the notion that the decisions to trust are rational and calculative. The individual differences relevant to this profile include the tit-for-tat strategy (Axelrod, 1984) and equity sensitivity (Huseman, Hatfield, & Miles, 1987). The social contextual factors important to this profile are exchange relationships (Clark & Mills, 1979) and cultures of dignity (Leung & Cohen, 2011). These variables all tend to facilitate perceptions of others' trustworthiness, increase vigilance toward violations and lower idiosyncratic credits, as well as promote relatively positive attributions of the violations after they perceive remorse from transgressors.

Research has suggested that the tit-for-tat strategy is effective in eliciting cooperation in social relationships (Oskamp, 1971; Pruitt & Carnevale, 1993). Tit-for-tat is a *contingent* strategy that matches the other party's prior behavior in relationships (Rubin & Brown, 1975). Individuals who adopt the tit-for-tat are concerned with mutual outcomes and avoidance of being exploited at the same time (Van Lange & Visser, 1999). In the beginning of a relationship, individuals with the tit-for-tat strategy always begin with trust and cooperation (Axelrod, 1984). In the absence of information indicated otherwise, they assume others are trustworthy. However, the use of the tit-for-tat strategy implies that individuals should be ready to retaliate immediately in the case of defection. In other words, these people should

be vigilant toward violations and allow very few idiosyncratic credits. Once the other party resumes cooperation, individuals who adopt the tit-for-tat strategy will choose to cooperate again. Because of this, those who adopts the tit-for-tat strategy are considered to be rational and maximizing (Guttman, 1996). They are quick to forgive (Hargreaves Heap & Varoufakis, 2004) and not expected to make negative attributions of the violations.

Another individual-difference factor that is particularly relevant to this profile is equity sensitivity. People vary in their degrees of sensitivity to equity (Fehr & Schmidt, 2000; Huseman et al., 1987). Specifically, individuals who are equity sensitive prefer a matching level of output, compared to their input, to that of others (Huseman et al., 1987). Preferences for equity and fairness have been found to motivate behaviors in situations that require cooperation (Croson, 1996; Karni, Salmon, & Sopher, 2008). We expect that those with high preferences for equity and fairness are more likely to adopt the tit-for-tat strategy in their relationship than others. During trust formation, concerns for similar mutual outcomes should facilitate trust building. Employees who prefer equity (comparable outcomes to contributions ratios between oneself and others) or underpayment (a smaller ratio compared to that of others) have been found to trust the organization more than those who prefer overpayment (a larger ratio compared to that of others; Kickul, Gundry, & Posig, 2005).

Equity sensitivity can also affect perceptions of whether a violation has occurred. Compared to those who prefer underpayment, individuals who prefer equity and overpayment have a lower threshold of violation perception (Morrison & Robinson, 1997). Further, when equity sensitive individuals perceive unfair treatment, they experience distress (Huseman et al., 1987). Because of these two factors, equity sensitivity is likely to increase vigilance toward potential violation that may compromise fairness. Trust dissolution should be rapid as a result. However, once the inequity is restored, such as by the repair efforts put forth by the trustee, trust restoration should be relatively fast. Perceptions of justice in repairing efforts have been found to increase positive attitudes toward the trustee (Gillespie & Dietz, 2009). Equity sensitive trustors should make positive attributions prompted by the trustee's attempts to rectify unfairness.

At the social contextual level, the tit-for-tat trust pattern can also be observed in exchange relationships. In exchange relationships, "benefits are given with the expectation of receiving a benefit in return" (Clark & Mills, 1979, p.12). Record-keeping, or concerns of individual inputs in joint tasks, are salient in exchange relationships (Clark, 1984). Consequently, the interaction logics of tit-for-tat are commonly used. During trust formation, individuals should be able to trust quickly, with the expectation that the other party will reciprocate the trust. However, idiosyncratic credits should be limited in an exchange relationship because of the strict rule of tit-for-tat and expectation of immediate reciprocity that is "in kind" with the exchange. Trustors in exchanges relationship are also vigilant toward deviations of the exchange norms (Clark & Chrisman, 1994). Trustors should thus respond quickly to violations. During trust restoration, trust is expected to build again quickly during trust restoration once the norms of exchange have been reestablished.
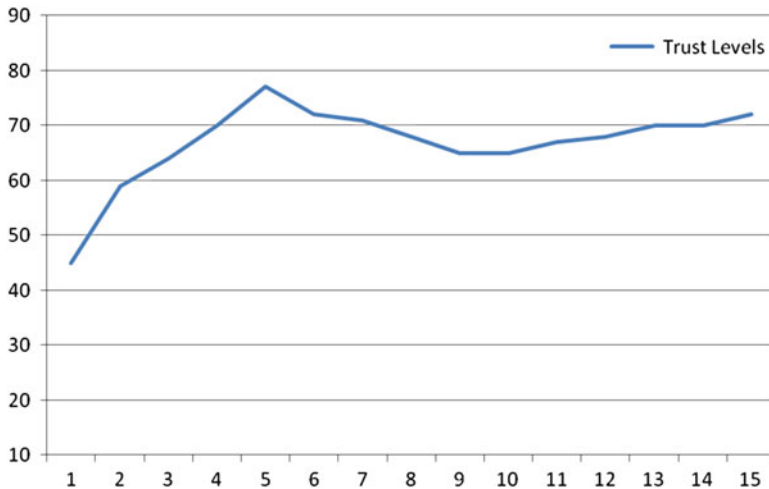
Fig. 4 The seizing and freezing trust profile with a fast formation, a slow dissolution, and a slow restoration

Culture can also act as a social contextual factor to influence this pattern. In particular, this pattern can be observed in the dignity cultures (Leung & Cohen, 2011). The culture of dignity is closely related to individualism which endorses the rational principles and individual rights (Gelfand, Bhawuk, Nishii, & Bechtold, 2004; Triandis, 1995). In dignity cultures, individuals' personal worth is intrinsic within each person and cannot be stolen by others. Dignity cultures thus value one's interval evaluation of the self and a moderate version of the tit-for-tat strategy in interacting with others (Leung & Cohen, 2011). People in dignity cultures generally have higher trust toward others because of the belief that dignity, or self-worth, is inherent in all human beings. They also believe that morality and integrity are attributes innate to people (Leung & Cohen, 2011; Ross, 1930). However, their concerns for equity and fairness should make them sensitive to trust violations and lead to fast trust dissolution. Moreover, the cultures of dignity endorse governing by internal standards rather than social sanctions (Kashima et al., 2004). Trustors in dignity cultures should therefore be less inclined to make negative attributions of the violations during trust restoration, believing that the trustee would experience guilt about the transgression (Leung & Cohen, 2011).

3.4 Seizing and Freezing Trust Profile: Fast Formation, Slow Dissolution, and Slow Restoration

Figure 4 shows the seizing and freezing trust profile, which is characterized by a relatively quick trust formation in the beginning of a trust relationship, a prolonged trust dissolution when violations have occurred, and a slow and difficult trust
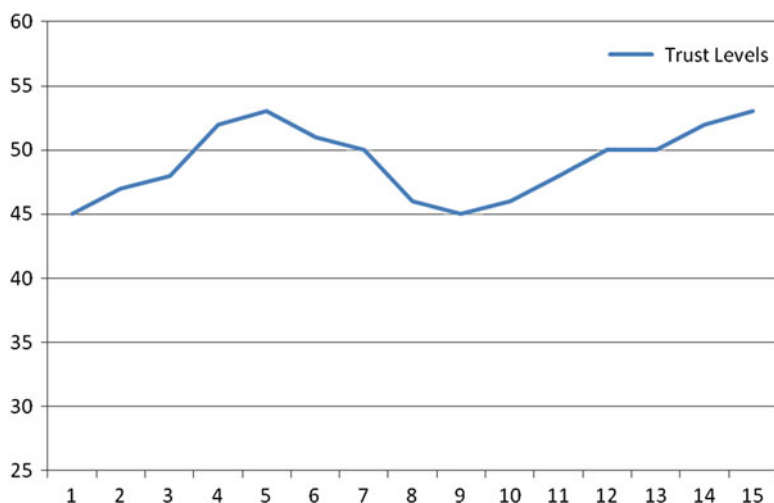
restoration after dissolution. Fundamental to this profile is the notion that trustors have a tendency to build trust quickly, but become affixed to the trust decisions made in the prior phase and slow to change their trust levels. This tendency for trustors' attitudes to stay close to a pre-established position is akin to the concept of attractors in dynamical modeling (Nowak & Vallacher, 1998). Attractors emerge in cognitive networks when individuals cannot fully integrate all available information and increasing rely on prior evaluative parameters as heuristics (Nowak & Vallacher).

An individual-difference variable that is highly relevant to this profile is need for closure (Kruglanski & Webster, 1996). The social contextual factors relevant to this profile are time pressure (De Dreu & Carnevale, 2003) and power distance with authority (Carl, Gupta, & Javidan, 2004). Other factors that can also affect this pattern are noise and fatigue (Kruglanski & Webster). These variables facilitate perceptions of others' trustworthiness, decrease vigilance toward violations and increase the number of idiosyncratic credits, but lead to negative attributions once the trustors eventually recognize the violations.

As an individual-differences dimension, need for closure (NFC) refers to "individuals' desire for a firm answer to a question and an aversion toward ambiguity" (Kruglanski & Webster, 1996). Trust building has been found to be more likely in certain versus uncertain situations (Goto, 1996) and it generally saves trustors time and effort to assume others' are trustworthiness in the beginning of relationships (Lewicki, Tomlinson, & Gillespie, 2006). Driven by their orientation for certainty, high NFC individuals are expected to trust quickly when they find evidence of the ability, integrity, and benevolence of the trustee. Based on the initial evidence of trustworthiness, trustors high on NFC would begin trust formation quickly to avoid prolonged uncertainty.

After seizing on a closure, NFC continues to motivate individuals to maintain the closure (Kruglanski & Webster, 1996). Therefore, high NFC individuals may be inclined to "freeze" on their trust in the trustee during trust dissolution. They may be less vigilant toward signs of trust violations as they focus on prior knowledge about the trustee and ignore subsequent information. They will also be more willing to give the trustees idiosyncratic credits because of the perceptions that trustees are trustworthy. However, once individuals with high NFC do decide to lower their trust, they should freeze on the notion that the trustee is untrustworthy and be disinclined to rebuild their trust in the trustee again. As these trustors freeze on the trust violations and the notion that trustors have acted in an untrustworthy manner, it would be difficult for them to make positive attributions of the violations and engage in trust restoration. Indeed, experimental work has demonstrated that people high on NFC exhibited abrupt conflict escalation, without a gradual progression, and difficulty in de-escalation (Bui-Wrzosinska, Cichock, Nowak, & Formanowic, 2009). The motivation to avoid ambiguity and maintain congruence promotes a nonlinear trust pattern of attractor dynamics of moving from one stable state to another (Vallacher & Nowak, 2007).

Related to dispositional NFC, we predict that social contextual factors that promote NFC should similarly lead to this trust pattern of fast formation, slow

dissolution, and slow restoration. For example, time pressure has been found to increase NFC (De Dreu & Carnevale, 2003) by increasing the use of heuristics in information processing and decision making (De Dreu & Carnevale, 2003; Jamieson & Zanna, 1989; Kruglanski & Freund, 1983). We thus expect that, under high time pressure, people will build trust quickly in trust formation, as trustors in general assume others' trustworthiness in the beginning of trust relationships (McKnight et al., 1998). Research has indeed found evidence of "swift trust" formation in temporary teams (Meyerson, Weick, & Kramer, 1996). Once perceptions of trustworthiness are formed, trustors under time pressure may rely on this belief heavily during trust dissolution. The heightened NFC of these trustors should lead them to be low on vigilance and allow more idiosyncratic credits than those trustors not under time pressure. However, when the trustors cannot ignore the signs of violations anymore, they will shift their position by changing their perceptions of the trustees from being trustworthy to untrustworthy. Time pressure will lead the trustors to "freeze" on this new position and the negative perceptions during trust restoration (Kruglanski & Webster, 1996), rendering the trust rebuilding process slow and difficult.

At the cultural level, this trust pattern may be found in cultures high on power distance, particularly with an authority. Power distance has been defined as "the degree to which members of an organization or society expect and agree that power should be shared unequally" (Carl et al., 2004, p. 517). In high power distance cultures, paternalistic leadership that demonstrates humane considerations and promotes hierarchical harmony is common and effective (Carl et al.). As a result, people in high power distance culture generally expect the authority figures to be high on ability, benevolence and integrity. Further, obedience and deference to authority is a key feature in high power distance cultures. People should thus be less vigilant toward violations from authorities and allow them more idiosyncratic credits. Individuals with high power, such as leaders, are generally given more latitude and idiosyncratic credits to violate group norms (Hollander, 1958; Homans, 1950). Greenberg and Alge (1998) predicted that subordinates tend to give violators of high power the benefit of the doubt, to avoid escalation of conflict. These effects should be stronger in high power distance cultures. However, because of the obligation and benevolence expected from authorities in high power distance cultures, trust restoration may be slow and difficult once the subordinate decides to lower their trust in the authorities. Brockner, Tyler, and Cooper-Schneider (1992) indeed showed that when subordinates who were highly committed to the authority ultimately found violation of fairness from the authority, their attitudes toward the authority became more negative than those of the subordinates low on commitment.

3.5   Assessment Trust Profile: Slow Formation, Slow Dissolution, and Slow Restoration

As can be seen in Fig. 5, an assessment trust profile is characterized by a prolonged trust formation in the beginning of a trust relationship, a prolonged trust dissolution

Fig. 5 The assessment trust profile with a slow formation, a slow dissolution, and a slow restoration

when violations have occurred, and a prolonged trust restoration after dissolution. Fundamental to this profile is the notion that the decision to trust in each phase should be a careful and slow process that cannot be easily made. Trustors would prefer to have as complete information about the trustee and the situation as possible and to weigh the pros and cons before making any trust decisions. The individual differences relevant to this profile include the assessment regulatory mode (Kruglanski et al., 2000) and need to avoid closure (Kruglanski & Webster, 1996). The social contextual factor relevant to this profile is public situations (Kruglanski & Webster, 1996). These variables lead trustors to resist making decisions about others' trustworthiness quickly, increase the likelihood of granting idiosyncratic credits to allow the benefit of the doubt, and refrain from making attributions of the violations immediately after violations.

The slow formation, dissolution, and restoration trust pattern can occur with individuals high on the regulatory mode of assessment. Assessors are concerned with "critically evaluating entities or states, such as goals or means, in relation to alternatives in order to judge relative quality" (i.e. judging the quality of something by considering both its merits and demerits in comparison with an alternative; Kruglanski et al., 2000, p.794). Individuals high on assessment prefer to have as much information as possible when evaluating situations and making decisions. Uncertainties increase the likelihood that trustors give the benefit of the doubt (Carson, Madhok, Varman, & John, 2003). Therefore they should be slow to trust because they need time and information to determine the trustee's trustworthiness. They are also slow to lower their trust because they are unlikely to jump to a conclusion at the first sight of a violation, as assessors evaluate both positive and negative information lengthily (Kruglanski et al., 2000). However, once their trust is

decreased, it will take an equal amount of time and efforts, if not more, to verify the trustee's trustworthiness, which makes them slow to trust again. Moreover, assessors are likely to make negative attributions of the violations during trust restoration and prolong the recovery process because assessment has been shown to have a positive correlation with social anxiety and a negative correlation with self-esteem (Higgins, Kruglanski, & Pierro, 2003).

Another individual difference that can promote the assessment trust pattern is the need to avoid closure. Need to avoid closure is on the opposite end of the continuum from need for closure (Kruglanski & Webster, 1996). Need to avoid closure therefore refers to individuals' desire to "suspend judgmental commitment" (Kruglanski & Mayseless, 1990, p.196). Because need to avoid closure is motivated by the benefits of suspending judgments and the costs of committing cognitive closure and even judgment mistakes (Freund, Kruglanski, & Shpitzajen, 1985; Webster, 1993), trustors high on need to avoid closure should spend a prolonged period of time to gather and consider the three facets (ability, integrity, and benevolence) related the trustee's trustworthiness. During trust dissolution, the trustors may grant the trustee high idiosyncratic credits. Individuals high on need to avoid cognitive closure are capable of generating alternative explanations (Mayseless & Kruglanski, 1987), and are thus more likely to give the benefit of the doubt when facing violations. Further, violations are sometimes ambiguous in their meanings, and trustors high on need to avoid closure would want to be certain about the situations before making a decision to lower trust. The process of trust dissolution would thus be slow. Need to avoid closure also affects how individuals make casual attributions (Webster). In trust restoration, these trustors should avoid making quick decisions about the intentionality, controllability, and stability of the violations.

At the social-contextual level, the assessment trust pattern may also be found when a trustor is making trust decisions in public, representing constituents. When the trustor is representing a group of constituents, the accountability involved and the public nature of the situation should lead the trustor to be slow to trust. Prior negotiation research has shown that constituents can lead their representative to take a contentious stand toward the opponent (Adams, 1976), particularly in individualistic cultures (Gelfand & Realo, 1999). Further, concerns for accuracy, such as those prompted by accountability, should promote prolonged decision-making (Kruglanski & Mayseless, 1990). Because of these factors, trust building should be slow when the trustor is responsible to a group of constituents.

Likewise, the representative may be reluctant to lower their trust when signs of violation appear in order to maintain an impression of positive outcome and avoid accountability. When a trust relationship can be observed in public, the potential costs and benefits associated with identifying trust violations should further heighten the desire for accurate trust decisions in the relationship. Trustors with the constituents should thus spend time to evaluate both favorable and unfavorable information about the trustee (Reckman & Goethals, 1973). Costly judgment mistakes in public situations have been found to reduce need for closure and suspend judgmental commitment (Kruglanski & Webster, 1996). Finally, trust
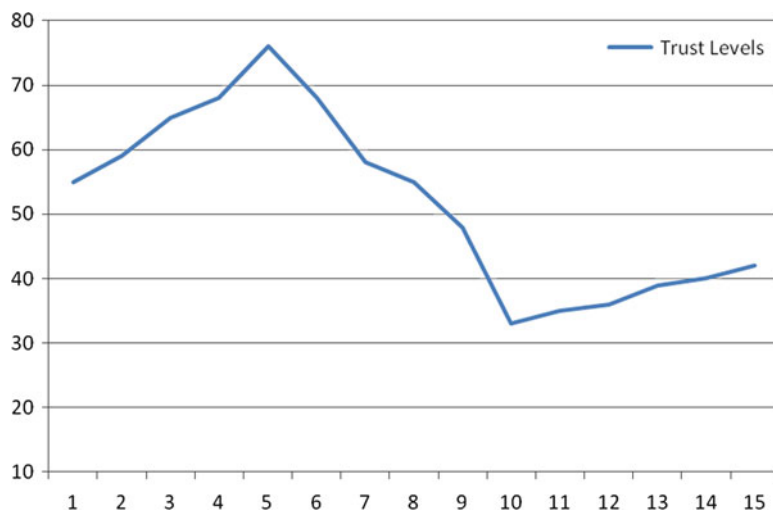
Fig. 6 The grim trigger trust profile with a fast formation, a fast dissolution, and a slow restoration

in the public situations and with constituents would be particularly difficult to repair after violations, as conflicts in this situation tend to be more severe than between individuals (Polzer, 1996). Additionally, because the trust decision will be made in public, the concern for accuracy should prolong the process of trust restoration.

3.6 Grim Trigger Trust Profile: Fast Formation, Fast Dissolution, and Slow Restoration

As Fig. 6 shows, a grim trigger trust profile is characterized by a quick trust formation in the beginning of a trust relationship, a quick trust dissolution when violations have occurred, and a slow and difficult trust restoration after dissolution. The name "grim trigger" reflects the fact that a single defection from the partner "triggers" one's defection in all subsequent interactions (Campbell, 2006). This profile is an epitome of "Trust, but verify," a catchphrase of President Ronald Reagan during the Cold War. Fundamental to this profile is the notion that trustors enter a trust relationship with positive perceptions of the trustees but expect a possibility of violations. More specifically, the grim trigger is a strategy used in relationships that engages in permanent retaliation once a violation has occurred (Axelrod, 2000). Individuals who adopt the grim trigger always cooperate in the absence of trust violations (Axelrod, 2000). Therefore, they should build trust quickly to sustain the cooperation (Hwang & Burgers, 1997). However, the trust relationship can only continue if no trust violation has occurred. Individuals with the grim trigger strategy would choose to stop the interactions for the slightest deviation from the relationship

norms (Friedman, 1971), even if it is only a single incident (Buskens & Weesie, 2000). Trustors with the grim trigger strategy therefore should have a high level of vigilance toward violations and grant practically no idiosyncratic credits to the trustees. Further, these trustors will not cooperate again once defection has been found (Friedman, 1971). The grim strategy is thus considered to be completely unforgiving (Axelrod, 2000). These trustors are expected to make very negative attributions about the violations and refuse to rebuild trust with the same trustees.

The individual differences relevant to this profile include interpersonal orientation (Rubin & Brown, 1975) and the social contextual factor relevant to this profile is performance orientation (Javidan, 2004). These variables promote positive perceptions of others' trustworthiness. However, they heighten vigilance toward violations, lower idiosyncratic credits, and produce negative attributions of the violations. For example, people high on interpersonal orientation are "responsive to the interpersonal aspects of his relationship with others. He is both interested in, and reactive to, variation in the other's behavior" (Rubin & Brown, 1975, p.158). In negotiation contexts, interpersonal oriented individuals have been found to behave cooperatively toward a cooperative opponent, but competitively toward a competitive opponent (Neu, Graham, & Gilly, 1988). Because they are attentive and responsive to the relational aspects of relationships (Graham, 1987) and are concerned about mutual benefits, individuals high on interpersonal orientation should exhibit fast trust building in the beginning of trust relationships. However, because interpersonally-oriented individuals have high sensitivity toward the behaviors of others (Rubin & Brown, 1975; Swap & Rubin, 1983), they are expected to exhibit high vigilance toward trust violations and grant few idiosyncratic credits. Once trust violations are identified, interpersonally-oriented individuals have high reactivity toward others' behaviors (Rubin & Brown, 1975) and will exhibit fast and negative responses in their trust. Indeed, individuals high on interpersonal orientation take the actions of another seriously (Rubin & Brown, 1975) and personally (Swap & Rubin, 1983). They are more likely to make internal attributions that those low on interpersonal orientation (Apostolon, Cotten, & White, 1981) and could thus attribute the violations as intentional, controllable, and a stable feature of the trustee, thus exhibiting a slow trust restoration.

At the social contextual level, performance cultures are relevant to the grim trigger trust profile, which have been defined as "the extent to which a human community encourages and rewards setting challenging goals, innovation, and performance improvement" (Javidan, 2004, p.276). High performance orientation creates a sense of urgency (Javidan, 2004) and individuals should thus be quick in their trust building. Indeed, a task-oriented culture has been theorized to have higher initial trust while a relationship-oriented culture should have lower initial trust (Schoorman, Mayer, & Davis, 2007). Because of the emphasis on contingent rewards, direct feedback, and competitiveness (Javidan, 2004), once a trust violation occurs, trustors in high performance orientation cultures should be responsive and lower their trust accordingly. In other words, we expect that the trustee has fewer idiosyncratic credits in high performance orientation cultures than low performance orientation cultures. Finally, trust restoration is expected to be slow in performance

oriented cultures. When facing failures, performance orientation can lead people to withdraw effort and commitment (Bell & Kozlowski, 2002). Further, performance orientation is related to the implicit theory of entity, believing that intelligence and ability are fixed attributes that cannot be changed (Button, Mathieu, & Zajac, 1996; Dweck & Leggett, 1988). These trustors are thus likely to use violations as diagnostic information about the trustee and make negative attributions of the violations in terms of the trustee's controllability and stability.

4   Discussion

Trust is critical for all forms of social relationships. It has been found to have ripple effects at multiple levels of analysis, including interpersonal (McAllister, 1995; Olekalns et al., 2002), group (Dirks, 1999; Jarvenpaa & Leidner, 1999), organizational (Graebner, 2009; Maguire & Phillips, 2008; Stahl & Sitkin, 2005), and national levels (Fukuyama, 1995; Putnam, 1993). Further, trust in relationships can evolve through multiple phases, from the initial development, through breakdown, to reestablishment. The growth and decline in trust across phases form trust dynamics that produce distinct patterns based on individual and social-contextual factors.

4.1   Theoretical Implications

Despite the fact that trust is dynamic, trust research has been criticized for neglecting the longitudinal nature of a trusting relationship by limiting it to one of the trust phases: trust building (Lewicki et al., 2006; Rousseau et al., 1998). Trust does not exist only in formation. Importantly, violations have been shown not to be exceptions but common occurrences (Jones & Burdette, 1994; Robinson & Rousseau, 1994). Multiple trust decisions are thus required throughout a trust relationship. Without taking into account the changes in the additional trust phases, our understanding of trust in relationships is incomplete. In this chapter, we took a holistic view by integrating trust trajectories in the three trust phases and examining how, together, they form divergent trust patterns that unfold over time in social relationships, which is of the first attempts to do so in the literature.

  While by no means exhaustive, this chapter has provided a coherent framework of trust dynamics through a simultaneous consideration of trust across the formation, dissolution, and restoration phases. In particular, we identified six commonly-occurring trust patterns: high trust, high distrust, tit-for-tat trust, seizing and freezing trust, assessment trust, and grim trigger trust. We also illustrated how these profiles are afforded by a multitude of individual difference and social contextual factors. The resulting framework provides a valuable springboard for

future research. Empirical research is needed to test this framework and identify additional individual difference and social-contextual factors that underlie trust trajectories.

4.2 Research Implications

Multiple methods are available to test the theory advanced in this chapter. For example, naturalistic methods could rely on experiential sampling of trust building, violation, and repair to test the ideas presented. Research can also use the critical incident technique to examine trust violations that occurred in the past and relate them to individual and social context factors. Interviews and content analysis of archival records are additional means to reveal how trust patterns unfold before and after violations over a period of time.

Experimental methods will likewise prove useful. For example, the proposed trust patterns can be examined in laboratory experiments using a variant of the Trust Game—the Investment Game (IG; Berg, Dickhaut, & McCabe, 1995). The IG is ideal for this research for a number of reasons. It affords social exchanges that mirror real-world interactions and permits observation of how violations and the resultant trust dissolution and restoration that occur during a trust relationship changes individuals' trust in their partner. In particular, an iterated IG (Cochard, Nguyen Van, & Willinger, 2004) is suitable for examining nonlinear and dynamic trust patterns because of the repeated measures of trust. We describe the structure of IG briefly below.

The IG typically involves two players, Player A and Player B. In this example, all participants are assigned to the role of Player A (trustor) and the computer-programmed partner is Player B (trustee). The game in this example consists of 15 rounds. The trust formation phase consists of rounds 1–5, and the trust dissolution phase consists of rounds 6–10. The remaining rounds (rounds 11–15) form the trust restoration phase. In the beginning of each round, Player A is given 100 coins and decides a proportion of the endowed coins (0–100) to entrust to Player B. This decision reveals how much Player A trusts Player B. The amount Player A sends to Player B is then tripled by the program, and Player B decides a portion of the tripled coins to return to Player A. Specifically, trust violations occur in the 6th, 7th, and 8th rounds, between the beginning and middle of the game. Violations that occur during rounds 6–8 allow trust to develop first. Violations occurring in the beginning of a relationship can lead to irreversible damage to trust (Lount, Zhong, Sivanathan, & Murnighan, 2008). Further, the three rounds of violations are designed so that participants would not perceive the violations as an isolated incident, which they may discount and thus keep trust unaffected (Sitkin & Roth, 1993). During these three violation rounds, the programmed partner keeps around 90% of the tripled coins with small random variation. In all other non-violation rounds, Player B returns approximately half of the tripled coins.

At the end of the 15th round, the game stops. As knowledge of the end of social exchanges tends to decrease cooperation (Murnighan, 1981), participants do not know how many rounds remain during the game before the end. This method can be readily combined with manipulation of the constructs discussed, such as high versus low self-esteem, differences in time pressure, surveillance and monitoring, third party and monitoring, as well as public versus private situation. Additionally, a series of measures on individual differences and social contextual factors can also be included to discern their effect on the dynamic trust patterns.

4.3 Practical Implications

In addition to the theoretical and research implications, the proposed framework has important practical implications. The different trust patterns help to identify when and at which phase trust management is particularly critical, depending on the individual and social-contextual factors. With this knowledge, third parties such as mediators can use this framework to intermediate conflicts. Pertinent individuals and situations can be targeted to manage violations and promote trust relationships. These trust patterns extend the current knowledge about trust behaviors. Our theorized trust dynamics across phases indicate that the seemingly divergent trust patterns individuals exhibit are governed by logics rational to the individuals and situations involved.

Finally, this chapter also focuses on the significant role that culture plays in trust dynamics. The trust field as a whole is limited by its lack of cross-cultural and intercultural research (Schoorman et al., 2007). Specifically, we identified the effects of five cultural dimensions—honor, dignity, collectivism, power distance, and performance orientation—on the trust phases. Our analysis suggests that trust patterns can at times vary dramatically across cultures. In the age of globalization (Ashkanasy, Härtel, & Daus, 2002), trust is increasingly important in determining our societal and global well-being. The proposed framework provides an initial roadmap to facilitate understanding about trust in the cross-cultural and intercultural context. The framework also provides guidance for individuals to utilize a variety of individual and social-contextual factors to manage trust in a particular cultural context.

5 Conclusion

In this chapter, we proposed a framework that examines six distinct profiles of trust trajectories across three trust phases. We further identified the individual and social contextual factors unique to each trust profile. Combined with rigorous and appropriate methodology that we recommended, this theory holds the potential

to provide researchers a holistic and dynamic understanding of trust in social relationships with implications to positively impact the field of trust research and the future of our society.

References

Adams, J. S. (1976). The structure and dynamics of behavior in organizational boundary roles. In M. D. Dunnette (Ed.), *Handbook of industrial organizational psychology* (pp. 1175–1199). Chicago: Rand McNally.

Andersson, L. M. (1996). Employee cynicism: An examination using a contract violation framework. *Human Relations, 49*, 1395–1418.

Andersson, L. M., & Bateman, T. (1997). Cynicism in the workplace: Some causes and effects. *Journal of Organizational Behavior, 18*(5), 449–469. Retrieved from Business Source Complete database.

Apostolon, M., Cotten, J., & White, M. (1981). *Causal attributions of high- and low-IO subjects*. Unpublished manuscript, Medford, MA.

Arnold, J. A., & O'Connor, K. M. (2006). How negotiator self-efficacy drives decisions to pursue mediation. *Journal of Applied Social Psychology, 36*, 2649–2669. doi:10.1111/j.0021-9029.2006.00121.x.

Ashkanasy, N., Härtel, C., & Daus, C. (2002). Diversity and emotion: The new frontiers in organizational behavior research. *Journal of Management, 28*(3), 307–338. doi:10.1177/014920630202800304.

Axelrod, R. (1984). The evolution of cooperation. New York: Basic Books.

Axelrod, R. (2000). On six advances in cooperation theory. *Analyse and Kritik, 22*, 130–151.

Bell, B. S., & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology, 87*, 497–505.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. doi:10.1006/game.1995.1027.

Beugré, C. (2005). Reacting aggressively to injustice at work: A cognitive stage model. *Journal of Business and Psychology, 20*(2), 291–301. doi:10.1007/s10869-005-8265-1.

Blough, D. S. (2001). Some contributions of signal detection theory to the analysis of stimulus control in animals. *Behavioural Processes, 54*(1–3), 127–136. doi:10.1016/S0376-6357(01)00154-1.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review, 98*, 294–310. doi:10.1257/aer.98.1.294.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). The requirements for trust in gulf and western countries. *Quarterly Journal of Economics, 125*, 811–828. doi:10.1162/qjec.2010.125.2.811.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk, and betrayal. *Journal of Economic Behavior and Organization, 55*(4), 467–484. doi:10.1016/j.jebo.2003.11.004.

Branzei, O., Vertinsky, I., & Camp, R. (2007). Culture-contingent signs of trust in emergent relationships. *Organizational Behavior and Human Decision Processes, 104*(1), 61–82. doi:10.1016/j.obhdp. 2006.11.002.

Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual Review of Psychology, 36*, 219–243. doi:10.1146/annurev.ps.36.020185.001251.

Brockner, J., Tyler, T. R., & Cooper-Schneider, R. (1992). The influence of prior commitment to an institution on reactions to perceived unfairness: The higher they are, the harder they fall. *Administrative Science Quarterly, 37*(2), 241–261. Retrieved from EBSCOhost.

Buchan, N. R., Croson, R. T. A., & Dawes, R. M. (2002). Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *The American Journal of Sociology, 108*(1), 168–206. Retrieved from JSTOR database.

Bui-Wrzosinska, L., Cichocka, A. K., Formanowicz, M. M., & Nowak, A. (2009, July). *Identifying trajectories of conflict: Presenting a research paradigm for the study of conflict reaction patterns over time.* Paper presented in the section Political Conflict, Violence, and Crisis for the 2009 ISPP Conference, Dublin, Ireland.

Burke, P. J., & Stets, J. E. (1999). Trust and commitment through self-verification. *Social Psychology Quarterly, 62*(4), 347–366. doi:10.2307/2695833.

Buskens, V., & Weesie, J. (2000). An experiment on the effects of embeddedness in trust situation: Buying a used car. *Rationality and Society, 12*(2), 227–253. doi:10.1177/104346300012002004.

Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes, 67*(1), 26–48.

Campbell, D. E. (2006). *Incentives: Motivation and the economics of information*. New York: Cambridge University Press.

Canevello, A., & Crocker, J. (2010). Creating good relationships: Responsiveness, relationship quality, and interpersonal goals. *Journal of Personality and Social Psychology, 99*(1), 78–106. doi:10.1037/a0018186.

Carl, D., Gupta, V., & Javidan, M. (2004). Power distance. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 513–563). Thousand Oaks, CA: Sage.

Carson, S. J., Madhok, A., Varman, R., & John, G. (2003). Information processing moderators of the effectiveness of trust-based governance in interfirm R&D collaboration. *Organization Science, 14*(1), 45–56. doi:10.1287/orsc.14.1.45.12811.

Cassidy, J. (1988). Child-mother attachment and the self in six-year-olds. *Child Development, 59*(1), 121–134. doi:10.2307/1130394.

Cialdini, R. B. (1996). Social influence and the triple tumor structure of organizational dishonesty. In D. M. Messick & A. E. Ten-Brunsel (Eds.), *Codes of conduct: Behavioral research into business ethics* (pp. 44–58). New York: Russell Sage.

Clark, M. (1984). Record keeping in two types of relationships. *Journal of Personality and Social Psychology, 47*(3), 549–557. doi:10.1037/0022-3514.47.3.549.

Clark, M., & Chrisman, K. (1994). Resource allocation in intimate relationships: Trying to make sense of a confusing literature. In *Entitlement and the affectional bond: Justice in close relationships* (pp. 65–88). New York: Plenum Press.

Clark, M., & Mills, J. (1979). Interpersonal attraction in exchange and communal relationships. *Journal of Personality and Social Psychology, 37*(1), 12–24. doi:10.1037/0022-3514.37.1.12.

Cochard, F., Van Nguyen, P., & Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior and Organization, 55*(1), 31–44. doi:10.1016/j.jebo.2003.07.004.

Cohen, D., & Hoshino-Browne, E. (2005). Insider and outsider perspectives on the self and social world. In R. M. Sorrentino, D. Cohen, J. M. Olson, & M. P. Zanna (Eds.), *Culture and social behavior* (pp. 49–76). Hillsdale, NJ: Erlbaum.

Collins, N., Ford, M., Guichard, A., & Allard, L. (2006). Working models of attachment and attribution processes in intimate relationships. *Personality and Social Psychology Bulletin, 32*(2), 201–219. doi:10.1177/0146167205280907.

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58*(4), 644–663. doi:10.1037/0022-3514.58.4.644.

Connell, J., Ferres, N., & Travaglione, T. (2003). Engendering trust in manager-subordinate relationships: Predictors and outcomes. *Personnel Review, 32*(5), 569–587. doi:10.1108/00483480310488342.

Cook, K. S., Yamagishi, T., Cheshire, C., Cooper, R., Matsuda, M., & Mashima, R. (2005). Trust building via risk taking: A cross-societal experiment. *Social Psychology Quarterly, 68*(2), 121–142. doi:10.1177/019027250506800202.

Costa, P., Zonderman, A., McCrae, R., & Williams, R. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology, 48*(4), 925–933. doi:10.1037/0022-3514.48.4.925.

Croson, R. (1996). Information in ultimatum games: An experimental study. *Journal of Economic Behavior and Organization, 30*(2), 197–213. Retrieved from EconLit database.

De Dreu, C. K. W., & Carnevale, P. J. (2003). Motivational bases of information processing and strategy in conflict and negotiation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 235–291). San Diego, CA: Elsevier Academic Press.

Dean, J., Jr., Brandes, P., & Dharwadkar, R. (1998). Organizational cynicism. *Academy of Management Review, 23*(2), 341–352. Retrieved from Business Source Complete database.

Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of Applied Psychology, 84*, 445–455. doi:10.1037/0021-9010.84.3.445.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87*(4), 611–628. doi:10.1037/0021-9010.87.4.611.

Dodge, K., Price, J., Bachorowski, J., & Newman, J. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology, 99*(4), 385–392. doi:10.1037/0021-843X.99.4.385.

Dorfman, P., Hanges, P., & Brodbeck, F. (2004). Leadership and culture variation: The identification of culturally endorsed leadership profiles. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Leadership, culture, and organizations: The GLOBE study of 62 societies* (pp. 669–719). Thousand Oaks, CA: Sage.

Downey, G., & Feldman, S. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology, 70*(6), 1327–1343. doi:10.1037/0022-3514.70.6.1327.

Doz, Y. L., & Hamel, G. (1998). *Alliance advantage: The art of creating value through partnering*. Cambridge, MA: Harvard University Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256–273.

Elangovan, A. R., & Shapiro, D. L. (1998). Betrayal of trust in organizations. *Academy of Management Review, 23*(3), 547–566. Retrieved from Business Source Premier database.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association, 7*, 235–266. doi:10.1162/JEEA.2009.7.2-3.235.

Fehr, E., & Schmidt, K. (2000). Fairness, incentives, and contractual choices. *European Economic Review, 44*, 1057–1068. doi:10.1016/S0014-2921(99)00046-X.

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*(5), 894–914.

Ferrin, D. L., Dirks, K. T., & Shah, P. P. (2006). Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of Applied Psychology, 91*(4), 870–883. doi:10.1037/0021-9010.91.4.870.

Fincham, F. D. (2000). Family violence: A challenge for behavior therapists. *Behavior Therapy, 31*(4), 685–693. doi:10.1016/S0005-7894(00)80038-6.

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology, 82*(6), 956–974. doi:10.1037//0022-3514.82.6.956.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. Reading, MA: Addison-Wesley.

Freund, T., Kruglanski, A. W., & Shpitzajzen, A. (1985). The freezing and unfreezing of impressional primacy: Effects of the need for structure and the fear of invalidity. *Personality and Social Psychology Bulletin, 11*(4), 479–487. doi:10.1177/0146167285114013.

Friedman, J. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies, 38*(113), 1–12. Retrieved from Business Source Complete database.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Gambetta, D. G. (1988). Can we trust trust? In D. G. Gambetta (Ed.), *Trust* (pp. 213–237). New York: Basil Blackwell.

Gelfand, M., Nishii, L., & Raver, J. (2006). On the nature and importance of cultural tightness-lossness. *Journal of Applied Psychology, 91*(6), 1225–1244. doi:10.1037/0021-9010.91.6.1225.

Gelfand, M. J., Bhawuk, D. P. S., Nishii, L. H., & Bechtold, D. J. (2004). Individualism and collectivism. In R. J. House, P. J. Hanges, P. J. M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations. The GLOBE study of 62 societies* (pp. 437–512). Thousand Oaks, CA: Sage.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., et al. (2011). *Differences between tight and loose societies: A 33-nation study*. Science, 332, 1100–1104.

Gelfand, M. J., & Realo, A. (1999). Individualism-collectivism and accountability in intergroup negotiations. *Journal of Applied Psychology, 84*(5), 721–736. doi:10.1037/0021-9010.84.5.721.

Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review, 34*(1), 127–145. Retrieved from PsycINFO database.

Gillespie, N. A., & Mann, L. (2004). Transformational leadership and shared values: The building blocks of trust. *Journal of Managerial Psychology, 19*(6), 588–607. doi:10.1108/02683940410551507.

Gilliland, S., Benson, L., & Schepers, D. (1998). A rejection threshold in justice evaluations: Effects on judgment and decision-making. *Organizational Behavior and Human Decision Processes, 76*(2), 113–131. doi:10.1006/obhd.1998.2801.

Glick, S., & Croson, R. (2001). Reputations in negotiation. In S. J. Hoch, H. C. Kunreuther, & R. E. Gunther (Eds.), *Wharton on making decisions* (pp. 177–186). New York: Wiley.

Glückler, J., & Armbrüster, T. (2003). Bridging uncertainty in management consulting: The mechanisms of trust and networked reputation. *Organization Studies, 24*(2), 269–297. doi:10.1177/0170840603024002346.

Goto, S. (1996). To trust or not to trust: Situational and dispositional determinants. *Social Behavior and Personality, 24*(2), 119–131. doi:10.2224/sbp. 1996.24.2.119.

Graebner, M. (2009). Caveat venditor: Trust asymmetries in acquisitions of entrepreneurial firms. *Academy of Management Journal, 52*(3), 435–472. Retrieved from PsycINFO database.

Graham, J. L. (1987). A theory of interorganizational negotiations. In J. N. Sheth (Ed.), *Research in marketing* (pp. 163–183). Greenwich, CT: JAI Press, Inc.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *The American Journal of Sociology, 91*, 481–510. doi:10.1086/228311.

Greenberg, J., & Alge, B. (1998). Aggressive reactions to workplace injustice. In *Dysfunctional behavior in organizations: Violent and deviant behavior* (pp. 83–117). Stamford, CT: Elsevier Science/JAI.

Guttman, J. M. (1996). Rational actors, tit-for-tat types, and the evolution of cooperation. *Journal of Economic Behavior & Organization, 29*(1), 27. Retrieved from EBSCOhost.

Hargreaves Heap, S., & Varoufakis, Y. (2004). *Game theory: A critical text* (2nd ed.). New York: Routledge.

Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52*(3), 511–524. doi:10.1037/0022-3514.52.3.511.

Heider, F. (1958). *The psychology of interpersonal relations*. Mahwah, NJ: Lawrence Erlbaum Associates.

Higgins, E., Kruglanski, A. W., & Pierro, A. (2003). Regulatory mode: Locomotion and assessment as distinct orientations. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 293–344). San Diego, CA: Elsevier Academic Press. doi:10.1016/S0065-2601(03)01005-0.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological Review, 65*(2), 117–127. doi:10.1037/h0042501.

Holmes, J., & Boon, S. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behaviour* (pp. 190–211). New York: Cambridge University Press.

Homans, G. C. (1950). *The human group*. New York: Harcourt/Brace.

Hovland, C. I., Harvey, O. J., & Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology, 55*(2), 244–252. doi:10.1037/h0048480.

Huseman, R., Hatfield, J., & Miles, E. (1987). A new perspective on equity theory: The equity sensitivity construct. *Academy of Management Review, 12*, 222–234. Retrieved from Business Source Complete database.

Hwang, P., & Burgers, W. P. (1997). Properties of trust: An analytical view. *Organizational Behavior and Human Decision Processes, 69*(1), 67–73. doi:10.1006/obhd.1996.2673.

Jamieson, D., & Zanna, M. (1989). Need for structure in attitude formation and expression. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 383–406). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791–815. Retrieved from Business Source Premier database.

Javidan, M. (2004). Performance orientation. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 437–512). Thousand Oaks, CA: Sage.

Jones, W. H., & Burdette, M. P. (1994). Betrayal in relationships. In A. L. Weber & J. H. Harvey (Eds.), *Perspective on close relationships* (pp. 243–262). Boston: Allyn & Bacon.

Jones, G., & George, J. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*, 531–546. doi:10.2307/259293.

Jung, D. I., & Avolio, B. J. (2000). Opening the black box: An experimental investigation of the mediating effects of trust and value congruence on transformational and transactional leadership. *Journal of Organizational Behavior, 21*(8), 949–964. Retrieved from Business Source Complete database.

Kanter, D., & Mirvis, P. (1989). *The cynical Americans: Living and working in an age of discontent and disillusion*. San Francisco: Jossey-Bass.

Karni, E., Salmon, T., & Sopher, B. (2008). Individual sense of fairness: An experimental study. *Experimental Economics, 11*(2), 174–189. doi:10.1007/s10683-007-9165-1.

Kashima, Y., Kashima, E., Farsides, T., Kim, U., Strack, F., Werth, L., et al. (2004). Culture and context-sensitive self. *Self and Identity, 3*, 125–141.

Kickul, J., Gundry, L., & Posig, M. (2005). Does trust matter? The relationship between equity sensitivity and perceived organizational justice. *Journal of Business Ethics, 56*(3), 205–218. Retrieved from Business Source Complete database.

Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The repair of trust: A dynamic bi-lateral perspective and multi-level conceptualization. *Academy of Management Review, 34*, 401–422. Retrieved from Business Source Premier database.

Kitayama, S., & Markus, H. (1999). Yin and yang of the Japanese self: The cultural psychology of personality coherence. In D. Cervone, Y. Shoda, D. Cervone, & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 242–302). New York: Guilford Press. Retrieved from EBSCOhost.

Koehler, J. J., & Gershoff, A. D. (2003). Betrayal aversion: When agents of protection become agents of harm. *Organizational Behavior and Human Decision Processes, 90*(2), 244–261. doi:10.1016/S0749-5978(02)00518-6.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50*, 569–598. doi:10.1146/annurev.psych.50.1.569.

Kramer, R. M., & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology, 46*(5), 1044–1057. doi:10.1037/0022-3514.46.5.1044.

Kruglanski, A., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19*(5), 448–468. doi:10.1016/0022-1031(83)90022-7.

Kruglanski, A., & Webster, D. (1996). Motivated closing of the mind: 'Seizing' and 'freezing'. *Psychological Review, 103*(2), 263–283. doi:10.1037/0033-295X.103.2.263.

Kruglanski, A. W. (1970). Attributing trustworthiness in supervisor-worker relations. *Journal of Experimental Social Psychology, 6*(2), 214–232. doi:10.1016/0022-1031(70)90088-0.

Kruglanski, A. W., & Mayseless, O. (1990). Classic and current social comparison research: Expanding the perspective. *Psychological Bulletin, 108*, 195–208.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 331–378). New York: Academic.

Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M. N., Pierro, A., Shah, J. Y., et al. (2000). To 'do the right thing' or to 'just do it': Locomotion and assessment as distinct self-regulatory imperatives. *Journal of Personality and Social Psychology, 79*(5), 793–815. doi:10.10371A1022-3514.79.5.793.

Kumashiro, M., Rusbult, C. E., Finkenauer, C., & Stocker, S. L. (2007). To think or to do: The impact of assessment and locomotion orientation on the Michelangelo phenomenon. *Journal of Social and Personal Relationships, 24*(4), 591–611. doi:10.1177/0265407507079261.

Leary, M., Tambor, E., Terdal, S., & Downs, D. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology, 68*(3), 518–530. doi:10.1037/0022-3514.68.3.518.

Leung, K., Bond, M., de Carrasquel, S., Muñoz, C., Hernández, M., Murakami, F., et al. (2002). Social axioms: The search for universal dimensions of general beliefs about how the world functions. *Journal of Cross-Cultural Psychology, 33*(3), 286–302. doi:10.1177/0022022102033003005.

Leung, A. K. Y., & Cohen, D. (2011). Within-and between-culture variation: individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of personality and social psychology*, 100, 507–526.

Lewicki, R., & Bunker, B. (1995). Trust in relationships: A model of development and decline. In M. Deutsch, B. B. Bunker, & J. Z. Rubin (Eds.), *Conflict, cooperation, and justice: Essays inspired by the work of Morton Deutsch* (pp. 133–173). San Francisco: Jossey-Bass.

Lewicki, R., & Bunker, B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114–139). Thousand Oaks, CA: Sage.

Lewicki, R., Tomlinson, E., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management, 32*(6), 991–1022. doi:10.1177/0149206306294405.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review, 23*(3), 438–458. doi:10.2307/259288.

Lount, R., Zhong, C., Sivanathan, N., & Murnighan, J. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin, 34*(12), 1601–1612. doi:10.1177/0146167208324512.

Maguire, S., & Phillips, N. (2008). 'Citibankers' at Citigroup: A study of the loss of institutional trust after a merger. *Journal of Management Studies, 45*(2), 372–401. Retrieved from PsycINFO database.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709–734. Retrieved from Business Source Premier database.

Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal, 48*(5), 874–888. Retrieved from EBSCOhost.

Mayseless, O., & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes*, 39, 162–183.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*, 24–59. Retrieved from Business Source Premier database.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology, 75*(6), 1586–1603. doi:10.1037/0022-3514.75.6.1586.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490. Retrieved from Business Source Premier database.

Meyerson, D., Weick, K., & Kramer, R. (1996). Swift trust and temporary groups. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166–195). Thousand Oaks, CA: Sage.

Mikulincer, M. (1998). Adult attachment style and affect regulation: Strategic variations in self-appraisals. *Journal of Personality and Social Psychology, 75*(2), 420–435. doi:10.1037/0022-3514.75.2.420.

Miles, R. E., & Creed, W. E. D. (1995). Organizational forms and managerial philosophies: A descriptive and analytical review. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 17, pp. 333–372). Greenwich, CT: JAI Press.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*(5), 961–978. doi:10.1037/0022-3514.46.5.961.

Miller, W. (1993). *Humiliation: And other essays on honor, social discomfort, and violence*. Ithaca, NY: Cornell University Press. Retrieved from EBSCOhost.

Mishra, A. (1993). Organizational responses to crisis: The role of mutual trust and top management teams. *Dissertation Abstracts International, 53*, 4398. Retrieved from PsycINFO database.

Morris, J. H., & Moberg, D. J. (1994). Work organizations as contexts for trust and betrayal. In T. R. Sarbin, R. M. Carey, & C. Eoyang (Eds.), *Citizen espionage: Studies in trust and betrayal* (pp. 163–187). Westport, CT: Praeger Publishers/Greenwood Publishing Group.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*(6), 949–971. Retrieved from EBSCOhost.

Morrison, E., & Robinson, S. (1997). When employees feel betrayed: A model of how psychological contract violation. *Academy of Management Review, 22*(1), 226–256. Retrieved from Business Source Complete database.

Murnighan, J. (1981). Defectors, vulnerability, and relative power: Some causes and effects of leaving a stable coalition. *Human Relations, 34*(7), 589–609. doi:10.1177/001872678103400706.

Murray, S. L., Holmes, J. G., & Collins, N. L. (2006). Optimizing assurance: The risk regulation system in relationships. *Psychological Bulletin, 132*(5), 641–666. doi:10.1037/0033-2909.132.5.641.

Neu, J., Graham, J., & Gilly, M. (1988). The influence of gender on behaviors and outcomes in a retail buyer-seller negotiation simulation. *Journal of Retailing, 64*(4), 427–451. Retrieved from PsycINFO database.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview.

Nowak, A., & Vallacher, R. (1998). *Dynamical social psychology*. New York: Guilford Press.

Olekalns, M., Lau, F., & Smith, P. L. (2002). *Spirals of trust: Identifying the factors that shape and sustain trust in negotiation*. Paper presented at the annual meeting of the International Association of Conflict Management Conference, Park City, UT. doi:10.2139/ssrn.305145

Oskamp, S. (1971). Effects of programmed strategies on cooperation in the prisoner's dilemma and other mixed-motive games. *Journal of Conflict Resolution, 15*(2), 225–259. Retrieved from JSTOR database.

Parasuraman, R. (1998). The attentive brain in aging and Alzheimer's disease. In R. R. Hoff-man (Ed.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 167–187). Washington, DC: American Psychological Association.

Pierro, A., Kruglanski, A. W., & Higgins, E. (2006). Regulatory mode and the joys of doing: Effects of 'locomotion' and 'assessment' on intrinsic and extrinsic task-motivation. *European Journal of Personality, 20*(5), 355–375. doi:10.1002/per.600.

Pitt-Rivers, J. (1966). Honour and social status. In J. Peristiany (Ed.), *Honour and shame* (pp. 19–78). London: Weidenfeld and Nicholson.

Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly, 1*, 107–142. doi:10.1016/1048-9843(90)90009-7.

Polzer, J. T. (1996). Intergroup negotiations: The effects of negotiating teams. *Journal of Conflict Resolution, 40*(4), 678–698. doi:10.1177/0022002796040004008.

Pruitt, D., & Carnevale, P. (1993). *Negotiation in social conflict*. Belmont, CA: Thomson Brooks/Cole Publishing.

Putnam, R. D. (1993). *Making democracy work: Civil traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*(5), 1281–1302. Retrieved from Business Source Premier database.

Reckman, R. F., & Goethals, G. R. (1973). Deviancy and group orientation as determinants of group composition preferences. *Sociometry, 36*(3), 419–423. doi:10.2307/2786343.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38*(2), 555–572. Retrieved from Business Source Premier database.

Robinson, S. L., & Rousseau, D. M. (1994). Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior, 15*(3), 245–259. doi:10.1002/job.4030150306.

Ross, W. D. (1930/2002). *The right and the good*. Oxford, UK: Oxford University Press.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal. *Journal of Personality, 35*(4), 651–665. doi:10.1111/j.1467-6494.1967.tb01454.x.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443–452. doi:10.1037/h0031464.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. Retrieved from Business Source Premier database.

Rubin, K., & Brown, I. (1975). A life-span look at person perception and its relationship to communicative interaction. *Journal of Gerontology, 30*(4), 461–468. Retrieved from PsycINFO database.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review, 32*(2), 344–354. Retrieved from Business Source Premier database.

Shah, J. Y., & Kruglanski, A. W. (2003). When opportunity knocks: Bottom-up priming of goals by means and its effects on self-regulation. *Journal of Personality and Social Psychology, 84*(6), 1109–1122. doi:10.1037/0022-3514.84.6.1109.

Shaver, P. R., & Mikulincer, M. (2003). The psychodynamics of social judgments: An attachment theory perspective. In J. P. Forgas, K. D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 85–114). New York: Cambridge University Press.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. Oxford, UK: Yale University Press.

Simons, T. L., & Peterson, R. S. (2000). Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *Journal of Applied Psychology, 85*(1), 102–111. doi:10.1037/0021-9010.85.1.102.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science, 4*(3), 367–392. Retrieved from Business Source Premier database.

Skinner, E., Edge, K., Altman, J., & Sherwood, H. (2003). Searching for the structure of coping: A review and critique of category systems for classifying ways of coping. *Psychological Bulletin, 129*(2), 216–269. doi:10.1037/0033-2909.129.2.216.

Stahl, G. K., & Sitkin, S. B. (2005). Trust in mergers and acquisitions. In G. K. Stahl & M. Mendenhall (Eds.), *Mergers and acquisitions: Managing culture and human resources* (pp. 82–108). Palo Alto, CA: Stanford Business Press.

Stewart, F. H. (1994). *Honor*. Chicago: Chicago University Press.

Strickland, L. H. (1958). Surveillance and trust. *Journal of Personality, 26*, 200–215. doi:10.1111/j.1467-6494.1958.tb01580.x.

Swap, W. C., & Rubin, J. Z. (1983). Measurement of interpersonal orientation. *Journal of Personality and Social Psychology, 44*(1), 208–219. doi:10.1037/0022-3514.44.1.208.

Tomlinson, E. C., & Mayer, R. C. (2009). The role of casual attribution dimensions in trust repair. *The Academy of Management Review, 34*(1), 85–104. Retrieved from PsycINFO database.

Triandis, H. (Ed.). (1976). *Variations in Black and white perceptions of the social environment*. Oxford, UK: U Illinois Press. Retrieved from EBSCOhost.

Triandis, H. (1995). *Individualism & collectivism*. Boulder, CO: Westview Press.

Triandis, H. C., McCusker, C., & Hui, C. H. (1990). Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology, 59*(5), 1006–1020.

Vallacher, R. R., & Nowak, A. (2007). Dynamical social psychology: Finding order in the flow of human experience. In A. W. Kruglanski, E. Higgins, A. W. Kruglanski, & E. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 734–758). New York: Guilford Press. Retrieved from EBSCOhost.

Valley, K. L., Moag, J., & Bazerman, M. H. (1998). "A matter of trust": Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization, 34*, 211–238. doi:10.1016/S0167-2681(97)00054-1.

Van Lange, P., & Visser, K. (1999). Locomotion in social dilemmas: How people adapt to cooperative, tit-for-tat, and noncooperative partners. *Journal of Personality and Social Psychology, 77*(4), 762–773. doi:10.1037/0022-3514.77.4.762.

Webster, D. M. (1993). Motivated augmentation and reduction of the overattribution bias. *Journal of Personality and Social Psychology, 65*(2), 261–271. doi:10.1037/0022-3514.65.2.261.

Weiner, B. (1986). Attribution, emotion, and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 281–312). New York: Guilford Press. Retrieved from PsycINFO database.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.

Wickens, C. D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). New York: Prentice Hall.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129–166. doi:10.1007/BF02249397.

Negotiating Cultural Conflicts Over Sacred Values

Kate Jassin, Hammad Sheikh, Nadine Obeid, Nichole Argo, and Jeremy Ginges

Abstract Most current approaches to negotiation of resource and political conflicts assume that parties to these conflicts are rational actors that weigh the costs and benefits of their choices, treat values as though they are fungible, and then act in a way that maximizes their benefits. However, recent research suggests that this is not the case. In other words, people do not treat all values as amenable to tradeoffs, but rather they distinguish between material values having to do with resource pricing and markets and sacred values that reside in the moral realm. Moreover, people seem to apply different reasoning to sacred vs. material values. Even more crucially, what is considered sacred and what is considered material varies among cultures. In this chapter we discuss research by us and others into the nature of sacred values in real world conflicts and the implications of the findings for ongoing political conflicts.

Keywords Sacred values • Material values • Negotiation • Political conflicts • Backfire effect

1 Introduction

Most current approaches to negotiation of resource and political conflicts assume that parties to these conflicts are rational actors that weigh the costs and benefits of their choices, treat values as though they are fungible, and then act in a way that maximizes their benefits (Atran, Medin, & Ross, 2005; Ginges, 1997; Varshney,

K. Jassin • H. Sheikh • N. Obeid • N. Argo
New School for Social Research, 80 Fifth Avenue, New York, NY 10011, USA

J. Ginges (✉)
Department of Psychology, New School for Social Research, 80 Fifth Avenue, New York, NY 10011, USA
e-mail: gingesj@newschool.edu

2003). This approach has the benefit of elegance, and is attractive because it suggests consistent means of negotiating conflicts across cultures and contexts. However, a growing body of research suggests that it is often inaccurate in at least three respects. First, people do not treat all values as fungible. Instead, they seem to distinguish between the everyday values of the marketplace (mundane values) and what are often referred to as sacred or protected values (Baron & Spranca, 1997; Tetlock, 2003). Second, people seem to apply different rules when making decisions about mundane versus sacred values (Ginges & Atran, 2009a, 2009b; Ginges, Atran, Medin, & Shikaki, 2007). As we will show, not only are these types of values not fungible, but standard negotiation strategies that assume fungibility (by offering material incentives or disincentives to induce compromise) will often backfire and increase resistance to compromise. Third, because sacred values are cultural products – the distinction between what is mundane and what is sacred varies widely across cultural contexts – a great deal of cultural knowledge and sensitivity is required when seeking to achieve cooperative outcomes in cross-cultural interactions. In this chapter we discuss some research carried out by our team and others into sacred values and the way they influence cultural and political conflicts. Specifically we will describe research into (1) the nature of sacred values in real world conflicts, (2) how current approaches to conflict negotiation may backfire when the conflict is over sacred values, (3) factors that may decrease opposition to compromising sacred values, and (4) the political implications of sacred values research and its application to ongoing political conflicts.

2   What Are Sacred Values?

> "The sacred and profane are always and everywhere conceived by the human intellect as separate genera, as two worlds with nothing in common . . . . They are different in kind . . . The mind experiences a deep repugnance about mingling"
>
>   –Emile Durkheim, The elementary forms of religious life

What does it mean for a value to be a "sacred value"? From one perspective the answer is simple and consistent across the literature. Most investigations into sacred values ask people to measure their commitments to different values along a material metric (e.g., how many jobs would be worth destroying an old-growth rainforest?). If people refuse to answer such questions, and they often do, they are said to be claiming sacred values – values protected from material tradeoffs (Baron & Spranca, 1997; Ginges et al., 2007; Tetlock, Kristel, Elson, Green, & Lerner, 2000). From another perspective there is disagreement. Some scholars interpret refusal to put a monetary value onto something as claims of infinite worth. This interpretation has led to the contention that sacred values are merely pseudosacred because infinite value is relatively easy to disprove. On logical grounds it is difficult to understand how people could have multiple infinite commitments to different values. Moreover, although people apparently have absolute commitment to sacred values when asked

how much money it might take for them to trade off such values (so called taboo tradeoffs), they will show a great deal more flexibility when asked to do things like tradeoff one sacred value for another (so called tragic tradeoffs; Baron & Leshner, 2000; Tetlock, 2003).

In our research we have interpreted refusals to measure sacred values along material metrics differently and more directly, arguing that such refusals should be taken at face value: as claims that some values occupy a domain separate from that of the marketplace and that these values are non-fungible with marketplace values. Emile Durkheim, in his classic work *The Elementary Forms of Religious Life* (1912/1954), argued that the sacred domain in religious life was defined by its very separation from the profane. Durkheim maintained that within each of these domains things can have relative value – just as some economic values are worth more than others so to some sacred things are more sacred than others. However, he suggested that the sacred and non-sacred cannot be valued along some common scale – and that any attempt to do so would result in a "deep repugnance".

> . . . the sacred and the profane are always and everywhere conceived by the human intellect as separate genre, as two worlds with nothing in common . . . They are different in kind (Durkheim, 1912/1995, p. 36).

While Durkheim deals with religion specifically, the tendency to partition the world into "sacred" and "profane" domains likely extends beyond religious life. There is a long tradition in the social sciences of making a distinction between the domains of the sacred and the profane (Iliev et al., 2009). Even the most non-religious believe that some things – the life of their child, love, old-growth rainforests – cannot be valued along some material metric. Both Mary Douglas (1966) and Marea Eliade (1957) suggest that our need to distinguish between sacred and non-sacred derives from a broader need to orientate ourselves in a messy, untidy and ambiguous universe. In the light of this discussion we can interpret refusals to value things along material metrics not as claims to infinite instrumental worth, but more directly as claims to the existence of different non-fungible domains that are bound by different and often incompatible rules (Ginges & Atran, 2009a, 2009b).

In this chapter we review research into sacred values that had two goals. First, we wished to test these two different conceptualizations of what it means to hold a sacred value. Second, we wished to investigate the way sacred values influence the trajectory of real world conflicts. To accomplish this we investigated the way people reason about compromising sacred values in experiments conducted in real world settings where our participants were key players in different disputes and for whom the sacred values of interest were significantly important to their lives.

3 The Backfire Effect

The notion that sacred values are "pseudosacred" seems to run counter to much historical, if anecdotal, evidence to the contrary: people undergoing severe deprivation and killing others in the name of abstract ideals such as nation, freedom, or

God (Ginges, 1997). A monk's self immolation, or a suicide bomber's sacrifice of life while killing others suggests that people will resist material incentives to compromise (Ginges & Atran, 2011; Ginges, Atran, Sachdeva, & Medin, 2011; Ginges, Hansen, & Norenzayan, 2009). However, most negotiation strategies assume that people make decisions based on some type of cost-benefit analysis. Thus, in order to increase the support for a proposed deal in a negotiation setting the benefits are increased, often by offering financial incentives for giving up important claims. A contemporary illustration is the case of the nuclear dispute between the U.S. and Iran. The U.S. has offered trade concessions to persuade Iran to give up its nuclear ambitions, which appears to be an emerging sacred value for Iran.

One initial goal of our research was to develop an experimental paradigm that could investigate the effect of material incentives to compromise over values considered sacred. In a series of experiments we first focused on the Israeli Palestinian conflict, running field experiments with Israeli settlers, Palestinian refugees, and Palestinian students who identified strongly with Hamas (Ginges et al., 2007).

Each was asked to give up something important to them in exchange for peace (the taboo deal). For example: 601 Israeli Settlers (who often believe that Eretz Ysrael–the territory roughly comprising Israel, the West bank, and Gaza–was promised to them by God) were asked to give up some of the land in favor for a two state solution; 535 Palestinian refugees were asked to compromise on the "right of return" to their hometowns and properties in Israel left in the war of 1967; and 719 Palestinian students were asked to compromise on their claim on East Jerusalem, which most Palestinians see as the capital of a future Palestinian state. Using a between subjects design, half the participants in each experiment were randomly assigned to a deal (see Table 1) that included added material compensation to "sweeten the deal" (of the taboo ± deal). The incentive was either personal (e.g. money for each Palestinian family) or collective (e.g. money to the Palestinian people).

None of our participants were agreeable to such deals. To determine whether the issues invoked sacred values we asked participants, in separate questions in a different part of the survey, whether they could imagine any compromise over these issues to bring some great gain to their people. Those who answered "no" were considered to be claiming a sacred value.

Importantly, the material offer backfired for those who considered the given issue sacred. These participants reacted with greater anger and disgust (i.e. moral outrage), and showed significant increases in support for violent opposition to the deals (e.g. by supporting bombing campaigns) when material compensations were offered for compromising on the sacred issue. In contrast, those participants who were against the deals, but for whom the issues were not sacred, responded in the opposite manner; for these participants material incentives significantly reduced moral outrage and support for violent opposition.

This backfire effect of material incentives to compromise sacred values is not restricted to the Palestinian-Israeli conflict and can occur even when the materially improved nature of the taboo + deal is transparent. For example in one study carried out in Indonesia, we surveyed a sample of students attending four different

Table 1 Sample scenarios of proposed Israeli and Palestinian trade-offs for peace

Sample	Deal type		
	Taboo deal	Taboo + deal	Symbolic deal
Israeli Settlers	Israel would recognize the historic and legitimate right of Palestinian refugees to return	The people of Israel would be able to live in peace and prosperity, free of the threat of war or terrorism	Palestinians would recognize the historic and legitimate right of the Jewish people to Eretz Israel
Palestinian refugees	Palestinians would recognize the historic and legitimate right of the Jewish people to Israel	The Palestinian people would be able to live in peace and prosperity, free of the threat of Israeli violence and aggression	Israel would recognize the historic legitimacy of the right of return
Palestinian students	Palestinians would recognize the sacred and historic right of the Jewish people to Israel	Israel will pay Palestine one billion US dollars a year for 10 years	Israel would recognize the historic legitimate right of the Palestinians to their own state and would apologize for all the wrongs done to the Palestinian people

Table 2 Sample taboo and taboo + deals presented to Indonesian sample

	Deals
Taboo	The United States and the European Union would agree to recognize the right of the Moslem Brotherhood to lead the government of Indonesia if elected in a free and fair manner, as long as there was agreement that Indonesia would not be ruled strictly according to sharia.
Taboo +	In return the United States and the European Union would give Indonesia a privileged trade agreement, resulting in substantial economic benefits to our people

madrassahs (Muslim boarding schools) that were associated with or founded by different elements of the Islamic political movement in that country (Ginges & Atran, 2009a). In the last few decades, Indonesia has seen a considerable degree of separatist and communal violence along ethnic and religious lines. Between 1990 and 2003, these intergroup conflicts claimed almost 10,000 casualties, and over one million people were displaced internally. One aspect of this conflict included militant Islamist groups fighting other factions with the goal of reorganizing Indonesian society following the rules of sharia (strict Islamic law).

We interviewed 102 students attending four different madrassahs Darussalam, Al-Husainy, Ibnu Mas'ud, and Al-Islam. All schools were associated with Islamic political movements. We asked the participants in our study to respond to different hypothetical deals (see Table 2), the first of which involved international recognition

for the right of the Moslem Brotherhood to lead Indonesia if they would give up the claim for rule by sharia. After assessing the participants' support for violent opposition against this deal, they were then offered the same deal with an additional material incentive (a privileged trade agreement with the United States and the European Union that would hugely benefit the Indonesian economy).

As in the case of the Israeli-Palestinian survey, the material incentive backfired. Participants rejected this offer more strongly than the offer without a material incentive, and showed higher support for violent opposition to the deal. These findings suggest that it is beneficial for negotiators to have a culturally sensitive understanding of issues under dispute. Commitment to very important but non-sacred issues appears to be flexible in the face of material incentives. However issues that are perceived to be sacred appear to obey a different logic. Here material incentives seem to violate a taboo against measuring commitment to sacred values along a material metric. For sacred values, material incentives can backfire. What then might negotiators do when cooperative outcomes require that one or both parties to a dispute compromise a sacred value?

4   The Symbolic Aspect of Intergroup Conflict

Although material incentives to compromise sacred values will often backfire, this should not be taken as meaning that sacred values can never be compromised. Along with anecdotal evidence showing peoples inflexibility over sacred values, there are many examples of compromise. To take one example, for many years in Israel there was a near unanimous consensus that it was taboo for Israelis to meet with the Palestine Liberation Organization, a taboo broken during the Oslo peace negotiations when Yitzach Rabin shook the hands of Yasir Arafat in 1993. One goal of our research is to investigate the contexts and mechanisms underlying compromise over sacred values.

Given the failure of material incentives to induce flexibility over sacred values, we began to focus on incentives that may not violate the taboo against measuring commitment to sacred values along material metrics. In experiments conducted in the West Bank and Gaza, we have examined whether opposition to peace deals involving compromise over sacred values is softened if the other side offers not material compensation but a strong symbolic concession over one of their own sacred values.

Using the matched samples described in the previous section (e.g. Israeli settlers, Palestinian refugees and Palestinian students) we examined opposition to two types of peace deals (see Table 1) in between-subjects experiments: the taboo deals previously described and the taboo deals plus symbolic concessions (symbolic deals) instead of material incentives. Symbolic concessions included recognizing the narrative of the other side, or symbolic renunciations of aspects of sacred values. The effect of these symbolic trade-offs (or gestures) was striking. Participants for

whom the deals invoked sacred values, whose response to deals hardened when offered material incentives, softened when offered symbolic concessions: in the symbolic condition participants showed significantly less opposition to the deals, less emotional outrage and less support for violence compared to those in the taboo condition. Significantly, the effect of symbolic effort appears mediated by reduced feelings of humiliation on the part of those asked to compromise their sacred values (Ginges & Atran, 2008). More generally, these results highlight the motivation of balanced agreements (Malhotra & Ginges, 2010). In the studies we have just described balanced means not so much the degree of compromise each side makes but refers instead to the domain of compromise.

It seems then that perceiving the other side as making a symbolic but difficult gesture may be an important breakthrough for successful negotiations in intractable conflicts. A symbolic gesture such as the recognition of the historic right of the Jewish people or recognition of the right of the Palestinians to their own state can take the negotiation process a step further towards reconciliation.

This discussion of the importance of symbolic aspects of intergroup conflicts and negotiations should not be taken as a minimization of the material aspect of many of these conflicts. Clearly, the resolution of intergroup conflict needs to account for material needs. However, dealing with the symbolic aspect of conflicts, including conflicts over sacred values, needs greater attention. Importantly, often those involved in difficult intergroup conflicts where religious and ethnic rights appear to collide often emphasize the symbolic.

To take one example, consider a study we ran with a sample of Lebanese Christians in the summer of 2007. Due to a growing Shiite and Sunni population and an increase in Christian emigration, Christians have found themselves a new minority in Lebanon (Obeid, 2010). Following the Israeli-Hezbollah July war of 2006, tension rose between Christian and Muslim sects living in Lebanon. Cultural and political tensions between the Maronite and Orthodox Christian denominations and the Shia and Sunni Muslim denominations respectively are pervasive in Lebanese society. They manifest in issues such as the power of Hezbollah, Muslim vs. Christian political representation, and the external influence of Iran, Syria and the U.S.

In this study we asked 60 Lebanese Christians, "What do you think the other communities need to do for conflicts within Lebanon to be resolved?" Most participants responded in terms of symbolic concessions rather than economic or specific needs (see Table 3). Specifically, 60% of responses pertained to recognition and acceptance of others (e.g., "acceptance of others", "tolerance for others", "communicating with others", "respect others", "open-mindedness"). Forty percent of responses pertained to loyalty to Lebanon over external allegiance or sectarianism (e.g., "interest of Lebanon first before Iran, Saudi, USA", "disconnect from regionalism", "separation from external countries, care for Lebanon first, be Lebanese first"). In contrast only 17% pertained to material aspects – such as the need to disarm militias or a fair share of power between sects.

Table 3 Typical symbolic and material concessions listed by Lebanese Christians in response to the question, "What do you think the other communities need to do for conflicts within Lebanon to be resolved?"

Symbolic concessions	Material concessions
Tolerance for others	Disarmament of Shia
Open-mindedness	Civic education systems
Respect for others	Fair share of power
Loyalty to Lebanon	Departure of Palestinians

5   Summary and Implications

These research findings suggest that it is necessary to obtain culturally sensitive understandings of the impact of sacred values on decision-making, particularly in negotiation settings. Mistakenly treating sacred values as very important regular values can potentially lead to significant problems: offering a material tradeoff in exchange for a sacred value can spur outrage, potentially derailing negotiations (e.g. the backfire effect). Importantly, sacred values can arise almost anywhere, as with the case of the nuclear issue in Iran, and abortion or even health care in the U.S.

Understanding which values within a conflict are sacred thus becomes necessary for negotiators. We are developing a locally adapted survey prototype to aid in the identification of relevant sacred values in specific contexts. The survey can be used to measure issues for which people will not make material tradeoffs; it may also seek to gauge whether parties to the conflict are aware of which values are sacred to the other side.

6   Implementation Challenges and Potential Solutions

We believe these studies add to a rich literature documenting significant cultural differences in how people represent the nature of conflict and negotiation contexts that influence expectations, goals, and biases (for reviews see Gelfand, Erez, & Aycan, 2007; Imai & Gelfand, 2009). Our work emphasizes the importance of understanding culturally specific representations of the nature of the issues under dispute. There are several ways that the findings herein can be useful to negotiators. Given the potential for the backfire effect, sacred values that are not central to the negotiation should be kept away from key negotiation points if at all possible. While this may seem obvious, it may be difficult. Negotiators may be mindful of mixing the sacred and profane during negotiations, but marginalized interest groups and factions hold incentives to deliberately try to mix the two; precisely to achieve backfire. The best strategy in such cases is to constantly bring discussion back to the hard material issues at hand.

When values identified as sacred are already entangled with the core issues, research suggests that they will best be dealt with via a non-material approach, or perhaps, a symbolic concession. As discussed in the previous section, research suggests that simply recognizing another stakeholder's sacred value–a form of symbolic concession–may pave the way for material discussion. Of course, such recognition has to be perceived as significant and valuable for both sides. In the Israel/Palestine studies described earlier, both sides have made clear that symbolic recognition of core values is a vital stepping stone to dispute resolution; thus, these particular symbolic concessions worked well in those contexts.

We recognize that getting polarized groups to recognize the most sacred values of the other side may be extraordinarily difficult from a political standpoint, requiring sensitive negotiation and mediating strategies within factions of each party to a conflict. It may even be the case that not recognizing a value known to be sacred to the other side is a sacred value for one of the parties–or its constituents. Moreover, policymakers may perceive symbolic recognition as a first step toward material concessions, and block the move.

As they are trying to foresee or manage such difficulties, it may be useful for negotiators to be aware of the political factors that make it more or less possible for stakeholders to block symbolic concessions. One critical factor is the internal stability and political structure of each stakeholder. These structural circumstances will affect how leadership proceeds at the negotiating table. For instance, consider a situation where multiple factions are competing for power within a weak state. For these factions, there is one primary means of "outbidding" their competitors: to attack the competition's "moderate" political moves as traitorous. In some contexts, recognizing the enemy can be considered taboo. For leaders in these situations, even if the majority of the population supports them, recognizing the pain or validity of the other side may mean political suicide. As we noted earlier, despite such difficulties leaders can be successful in breaking such taboos – the Israeli recognition of the Palestine Liberation Organization during the 1990s being one such example.

7   Future Research

Sacred values present an exciting direction for future research. Given that the findings discussed within this chapter are relatively new, many questions remain unanswered.

For instance, what is the nature of a sacred value? We know that sacred values have different attributes than material values, but what does that mean? Are they processed in the same ways that moral intuitions are, or do they have their own rules of activation? Similarly, when we say that sacred values are "protected from tradeoffs" (in contrast to material values), it seems to imply that they cannot have relative value. And yet, as we have argued previously in this chapter, just because

they operate within a different domain does not mean they cannot be ranked. Our, survey research in Israel and the Palestinian Territories has begun to show that such rankings exist. What are the operating rules for these values?

Given the tremendous implications of sacred values, two questions hold special urgency for future research. First, what causes sacred values and how do they form? With the case of Iran (Dehghani et al., 2009), we observed two important things: first, sacred values can form rather quickly; second, they appear to be prompted as much by outgroup resistance as by the issue itself. In Iran, popular support for the nuclear issue appeared to take on SV properties for some elements of the population within a matter of months. Factors leading to SV formation might have included any of the following (the list is not exhaustive): narratives that link resource issues to matters of identity; insecurity related to resource constraints, compounded by threats from the outside; a pervasive sense of impaired collective esteem within a people or nation following a defeat, broadly defined; or any social conflict where one group feels less powerful than another–perhaps inequity in the status quo encourages one group to conceptualize the conflict on a moral plane.

Second, sacred values can also be defused. As discussed earlier in this chapter, promising new research points to the possibility that an adversary's recognition of a sacred value–via acknowledgment or apology–may help return negotiation to a more instrumental plane.

References

Atran, S., Medin, D. L., & Ross, N. O. (2005). The cultural mind: Environmental decision-making and cultural modeling within and across populations. *Psychological Review, 112*(4), 744–776.

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied, 6*, 183–194.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16.

Dehghani, M., Iliev, R., Sachdeva, S., Atran, S., Ginges, J., & Medin, D. L. (2009). Emerging sacred values: Iran's nuclear program. *Journal of Judgment and Decision Making, 4*, 930–933.

Douglas, M. (1966). *Purity and danger*. New York: Routledge.

Durkheim, E. (1995/1912). *The elementary forms of religious life*. New York: Free Press.

Eliade, M. (1957). *The sacred and the profane*. New York: Harcourt Press.

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58*, 479–514.

Ginges, J. (1997). Deterring the terrorist: A psychological evaluation of different strategies for deterring terrorism. *Terrorism and Political Violence, 9*, 170–185.

Ginges, J., & Atran, S. (2008). Humiliation and the inertia effect: Implications for understanding violence and compromise in intractable intergroup conflicts. *Journal of Cognition and Culture, 8*, 281–294.

Ginges, J., & Atran, S. (2009a). Noninstrumental reasoning over sacred values: An Indonesian case study. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Moral judgment and decision making, Vol. 50). San Diego, CA: Academic.

Ginges, J., & Atran, S. (2009b). What motivates participation in violent political action: Selective incentives or parochial altruism? *Annals of the New York Academy of Sciences, 1167*, 115–123.

Ginges, J., & Atran, S. (2011). War as a moral imperative (not just practical policy by other means). In *Proceedings of the Royal Society*: *Biological Sciences*. Online first publication, February 16, 2011.

Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences, 104*, 7357–7360.

Ginges, J., Atran, S., Sachdeva, S., & Medin, D. (2011). Psychology out of the laboratory: The challenge of violent extremism. *American Psychologist, 66*(6), 507–519.

Ginges, J., Hansen, I. G., & Norenzayan, A. (2009). Religion and popular support for suicide attacks. *Psychological Science, 20*, 224–230.

Iliev, R., Sachdeva, S., Bartels, D. M., Joseph, C., Suzuki, S., & Medin, D. L. (2009). Attending to moral values. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Moral judgment and decision making, Vol. 50). San Diego, CA: Academic.

Imai, L., & Gelfand, M. J. (2009). Culture and negotiation: Interdisciplinary perspectives. In R. S. Bhagat & R. M. Steers (Eds.), *Handbook of culture, organizations, and work*. Cambridge, UK: Cambridge University Press.

Malhotra, D., & Ginges, J. (2010). Preferring balanced vs. advantageous peace agreements: A study of Israeli attitudes towards a two state solution. *Judgment and Decision Making, 5*, 420–427.

Obeid, M. (2010). A lebanese confession: Why religious politics is bad for Lebanon. *Harvard Kennedy School Review, 10*, 104–108.

Tetlock, P. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*, 320–324.

Tetlock, P. E., Kristel, O., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counter-factuals. *Journal of Personality and Social Psychology, 78*, 853–870.

Varshney, A. (2003). Nationalism, ethnic conflict and rationality. *Perspectives on Politics, 1*, 85–99.

Studying Trajectories of Conflict Escalation

Lan Bui-Wrzosinska, Michele Gelfand, Andrzej Nowak, and Laura Severance

Abstract The present chapter describes research in progress which is developing a simple, replicable methodology aimed at identifying the regularities and specificity of human behavior in conflict escalation and de-escalation processes. These research efforts will ultimately be used to study conflict dynamics across cultures. The experimental data collected through this methodology, together with case-studies, and aggregated, time-series macro data are key for identifying relevant parameters, systems' properties, and micro-mechanisms defining the behavior of naturally occurring conflict escalation and de-escalation dynamics. This, in turn, is critical for the development of realistic, empirically supported computational models. The article outlines the theoretical assumptions of *Dynamical Systems Theory* with regard to conflict dynamics, with an emphasis on the process of conflict escalation and de-escalation. Work on a methodology for the empirical study of escalation processes from a DST perspective is outlined. Specifically, the development of a progressive scenario methodology designed to map escalation sequences, together with an example of a preliminary study based on the proposed research paradigm, is presented. Implications of the approach for the study of culture are discussed.

Keywords Conflict escalation and de-escalation • Dynamical processes • Time • Cross-cultural differences

L. Bui-Wrzosinska (✉)
Warsaw School of Social Sciences and Humanities, Warsaw, Poland
e-mail: lan@iccc.edu.pl

M. Gelfand • L. Severance
Department of Psychology, University of Maryland, College Park, MD, USA
e-mail: mgelfand@umd.edu; leseverance@gmail.com

A. Nowak
Department of Psychology, Florida Atlantic University, Boca Raton, FL, USA
e-mail: anowak@fau.edu

1 Introduction

Understanding, predicting, and managing conflict are arguably among the most important challenges facing mankind. With increasing interdependence, the wellbeing of societies and their potential for growth and cooperation primarily depends on the way the global community is able to handle existing as well as emerging social conflicts. However, contemporary conflicts are constantly changing and evolving in a broad cultural context. In fact, conflict is a universal phenomenon, yet the way disputes are manifested can vary dramatically across cultures (Gelfand, Erez, & Aycan, 2007; Imai & Gelfand, 2010).

Nevertheless, cross-cultural research on conflict dynamics is in its infancy. While main effects of cultural dimensions, individual conflict styles, or utility functions on people's behavior are being widely investigated, and new, qualitative dimensions of culture are being explored, still little research has been studying the way culture impacts how people's behaviors evolve over time in conflict. However, in order to formulate accurate cross-cultural models and predictions on disputes dynamics in the interpersonal, organizational, social, or military context, we need to gather consistent data on the dynamic variance of conflict across cultures. In fact, some conflicts evolve in a gradual manner, while other rapidly burst into violence; a given behavior may be seen as mundane in one culture, while it may be understood as a major violation, and be a trigger leading to series of unexpectedly violent reactions in another culture. Conflict, from this perspective, presents a challenge to social sciences, as it requires new tools to help understand, predict, and manage the constantly evolving, dynamical character of the phenomenon.

Recently, a paradigm widely used in other areas of science – the dynamical systems approach (DST) – has been applied to the study of social conflict (Coleman, Bui-Wrzosinska, Vallacher, & Nowak, 2007; Coleman, Vallacher, Nowak, Bui-Wrzosinska, 2007; Vallacher, Coleman, Nowak, & Bui-Wrzosinska, 2010; Vallacher, Nowak, Bui-Wrzosinska, & Coleman, 2006). This theoretical advance has opened new avenues for the study of complex systems of conflict, bringing computational models, computer simulations and advanced conceptual tools to bear on studying conflict (see for examples, Guastello, Koopmans, & Pincus, 2008; Nowak & Vallacher, 1998; Vallacher & Nowak, 2007). Computational models and simulations hold the potential to advance not only the understanding of the dynamic interplay between culture and conflict, but also to have predictive value, which is critical for applications. However, empirical data is also crucial to ensure that computational models are based on realistic, tested assumptions and thus show predictive power. Such empirical data should utilize multiple methods of data collection, including case studies analysis, statistical, time-series data, fieldwork, and experimental psychological data allowing for causal inference.

The present chapter describes a work in progress on the latter effort – experimental psychological data. The main focus of the present project is the cultural context of conflict escalation and de-escalation dynamics. Specifically, we report our efforts toward the development of a simple, replicable methodology aimed at

identifying the regularities and specificity of conflict escalation and de-escalation patterns, which ultimately can be used to study conflict dynamics across cultures. We believe that experimental data collected through this kind of methodology can help identify relevant cultural parameters, and their effects on naturally occurring conflict escalation and de-escalation processes. This, in turn, can facilitate the development of realistic computational models built on empirically supported rules.

In what follows, we first outline the theoretical assumptions of DST with regard to conflict dynamics, with an emphasis on the process of conflict escalation and de-escalation. Next, the work in progress on a methodology for the empirical study of escalation processes from a DST perspective will be outlined. Specifically, the development of a progressive scenario methodology designed to track escalation sequences, together with an example of a preliminary study based on the proposed research paradigm, will be presented. Implications of the approach and methodological tools for the study of culture are then discussed.

2   The Dynamical Systems Approach to Conflict

Along with the dynamical systems approach to social psychology (Nowak & Lewenstein, 1994; Nowak & Vallacher, 1998; Vallacher & Nowak, 2007) social phenomena can be described with the use of some core Dynamical Systems Theory concepts. The term dynamical system is generally used to describe numbers of interconnected elements that change and evolve *over time*. From this perspective, for example, boiling water is viewed as the current state of a system of interacting molecules, the brain as a system of interacting neurons, or the society as a system of interacting individuals. A dynamical system can generally be conceptualized as the state of its elements at a given time. A system's behavior is a sequence of such states. To describe sequences of states, we need to identify key variables and parameters capturing the evolving characteristics of the system. Although key parameters such as temperature and density for instance, describing a system of interacting molecules in a state of steaming water, ice, or vapor seem relatively simple to measure and identify, specifying key parameters for the description of human systems undergoing different phases of conflict escalation still poses an important challenge to social sciences.

In existing DST methodology, the temporal evolution of a social interaction can be plotted on a state space, where the x axis represents one side of the interaction, and the y axis represents the other side. In their marital interactions studies, for example, Gottman, and colleagues (2002) have been analyzing and predicting the trajectory of marital interaction by tracking the state of the husband and the state of the wife, and computing the rules governing their state trajectories. For example, two people may be strongly coupled and the dynamics of the husband can be majorly predicted by the state of the wife, or conversely the states of the couple may be weakly linked, and evolve in a less predictable way. Additionally, if a given region of the state space is frequently visited, it can be predicted that it will stabilize over

time and govern the couple's dynamics. Developmental psychologists have likewise advanced a tool – *Gridware*, where the trajectory of parent/child interactions are mapped on a two dimensional grid (Granic & Lamey, 2002; Lewis, 2000; Lewis, Lamey, & Douglas, 1999). The state of the parent is observed, coded, and plotted on one axis and the state of the child on another axis. Such data allow us to identify the regions of such grid on which the states of the interaction tends to stabilize, and toward which the system will return after being perturbed.

Here, we propose that trajectories of conflict escalation and de-escalation cannot only be mapped on a grid, basing on observation data, but also experimentally tested as a sequence of one party's reaction to another party's conflict provoking behavior. Using this strategy, we can track the variance in whole patterns of responses for the same scenario of conflict provocation across different social, cultural, and psychological conditions: people can escalate gradually in response to gradual intensification of aggressive behaviors from the other party, but the same conditions can also lead to exaggerated response or, conversely, resistance to change and stabilization at a given level of intensity. Responses trajectories can also progress along some repeated cycles, or have unpredictable, irregular character.

One way to formally portray and systematize such results is to describe the dynamical properties of conflict escalation trajectories as attractor's dynamics. Generally, the dynamical systems approach to social psychology identifies and describes attractors in social systems as regions, toward which trajectories in a state space converge with time. It is common to distinguish among four classes of attractors (Eckmann & Ruelle, 1985; Schuster, 1984): *fixed-point*, *periodic*, *quasiperiodic*, and *chaotic*. Here, we discuss in particular fixed-point attractors. The method we use to assess attractor dynamics is to actively perturb the system through a sequence of conflict provocation stimuli. If a single, fixed-point attractor exists, the system will return to the same state after some time, thus one party's response will return to the same level of conflict intensity, regardless of the influence from the other party. In the case of multiple fixed-points, small perturbations will result in the system returning to its original state, but further changes of the control parameter may result in the system moving toward a different equilibrium: threshold effects are to be expected in the responses patterns. In dynamical social psychological terms similar dynamics have been understood as catastrophic scenarios of change (Tesser & Achee, 1994) and will be referred to as catastrophic (as opposed to gradual, incremental) escalation. Properties of such scenarios are of particular relevance for de-escalation and practical applications: the hysteresis effect described in catastrophe theory (Thom, 1975) for instance, explains how crossing certain thresholds in conflict escalation leads to irreversible changes, undermining the potential for further de-escalation.

Although theories of conflict, disputes, and especially conflict escalation generally address the dynamical character of conflict processes, empirical studies rarely follow the dynamical paradigm. Research has mainly been static, with emphasis on conflict styles, strategies, or stable, individual differences. Here, we propose that conflict escalation and de-escalation processes can be mapped, and analyzed with regard to their dynamical properties (rate of change, stabilization, trajectories).

Although emerging in the field of conflict studies, dynamical methodologies are effectively applied in different areas of psychological research (cf. Gottman, Swanson, & Swanson, 2002; Granic & Lamey, 2002; Levenson & Gottman, 1983; Lewis, 2000; Lewis et al., 1999). Dynamical methodologies involve observational data, where individuals, dyads or teams reaction patterns are tracked at different levels, from coded behavioral responses, through emotional states, to physiological or neural imaging.

Key in the formation of people's patterns of social interactions is the social and cultural context. Culture develops, limits, and expands people's behavioral repertoire, that is the range of the possible actions and reactions people consider in response to social situations. Below we discuss the initial development of a tool – the progressive scenario methodology – a work in progress toward experimental data collection aimed at testing the role of cultural parameters on changes of people's psychological repertoire in the course of conflict escalation and de-escalation.

2.1   Progressive Scenario Tool Development

The progressive scenario tool is mapping the scope of one party's behavioral options in response to another party's conflict provocation behavior. The manipulated parameter, representing the stimulus responsible for changes over time, is a linear progression of the other party's conflict provocation behavior. Through the use of this tool, numerous independent variables such as cultural factors or individual differences can be manipulated in order to track their effect on the course of escalation/de-escalation. Below, we describe the two components of the tool: (1) the stimuli and (2) the response scale.

The stimuli consist of a series of short descriptions (vignettes) of gradually escalating and de-escalating conflict behaviors displayed by a colleague at work in a situation of task interdependence ("you are working on a common project at work"). Fourteen subsequent vignettes are scaled according to the level of destructiveness and aggressiveness of the behavior they represent: the first seven scenes outline a scenario of progressive escalation of provocation by a colleague at work, from a relatively mild disagreement ("Your colleague criticizes your work") to open confrontation and humiliation ("During a company picnic, your colleague insults your partner/relative publicly"). The remaining seven scenes outline a progressive de-escalation scenario, with descriptions of conciliatory behaviors aimed at reversing each escalatory step ("Your colleague apologizes publicly for his inappropriate behavior toward your partner/relative").

The response scale includes a list of 30 behaviors, which maps a general behavioral repertoire available to people across different conflict situations, ranging from friendly dispute to extremely hostile and aggressive interactions. This repertoire was first scaled with regard to the level of destructiveness to the relationship between the parties each possible response represents, from relatively constructive ("talking it over") to extremely hostile and destructive behaviors ("hurting him/her as much as

possible"). However, further ongoing studies, including multidimensional scaling procedures across cultures reveal additional dimensions such as individual/social character of the responses, the honor dimension, as well as the verbal/physical dimension.

Items were generated via focus groups conducted with individuals working in organizations as well as discussions with subject matter experts (professional mediators, and scholars from the conflict resolution field). Respondents generated an extensive list (150 items) of all the possible reactions people can have in conflict situations. Items were then analyzed, and narrowed down to 30 items, with careful conservation of the wide spectrum (space) of possible behaviors.

Large samples of individuals scaled items along conflict dimensions. These efforts were aimed at collecting qualitatively informed items (focus groups and subject matter experts), with the possibility to translate qualitative properties to quantitative data (scaling of the items along social psychological dimensions), and thus map a party's response trajectory on the defined phase space with some relative precision given the qualitative character of the data. Currently, a complete scale has been designed to measure changes in the destructiveness of the response behavior. However, other order parameters along which the behavioral repertoire can be described are currently been included, such as degree of honor violation, individual/social dimension or trust.

2.2 Conflict Trajectories

Results from the questionnaire (responses on the scale of possible behaviors for each level of the provocation) can be mapped on a two dimensional space. The space is defined by the level of conflict provocation displayed by the other party (control parameter), and by the level of response destructiveness and aggression (order parameter). Results can be mapped as (1) a state trajectory, where the averaged point of response destructiveness is plotted for each step of provocation. (2) The "map of possible behaviors" (the ensemble of behaviors considered as possible to display at a given time) in response to a given level of conflict provocation from the other party. The response is then not a point on the scale, but rather a space, where different regions are loaded with concrete values for groups or individuals. This allow for the identification of patterns of most chosen behaviors, but also latent clusters of possible behavior that remain stable across situations. Results from studies using the described questionnaire constitute a starting point for modeling, causal inference, as well as for the testing of the effect of various parameters on a laboratory simulation of escalation dynamics.

To illustrate the possibilities of the scenario tool, we conducted a preliminary study investigating the impact of relational closeness between two parties on conflict escalation and de-escalation trajectories. The most obvious difference between close and more distant relationships is the valence of the relation between parties, biased toward more positive for close relationships. Another difference, less trivial and

Fig. 1 Trajectories of escalation for close and distant relationships
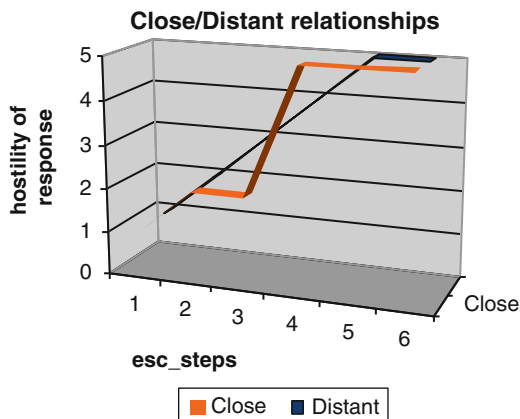
critical from the point of view of this study, is that close relationships exert pressure on the relation toward consistency in attitudes, emotions, and in the overall patterns of interactions between the parties, thus restraining the scope of the behavioral repertoire to positive behaviors, while more distant relationships induce fewer constraints on the relation.

The study revealed that closeness is related to an initial narrow behavioral repertoire, involving a set of unequivocally positive possible behaviors. Results show, however, that in response to stronger perturbations manipulated through conflict provocations, such systems react abruptly, with qualitative shifts in the relationship, while more distant relationships are characterized by gradual escalation trajectories. In fact, close relationships lead to trajectories with a major shift after a series of responses, where, despite contentious behaviors from the other party (from situations, where your friend does not answer your phone calls or emails, to situations, where he criticizes your work in front of other people), the responses initially remained at a very low level of hostility (the behavioral repertoire initially comprised: "listening to him/her, turning it agreeably into humor, etc . . . "). However, when a threshold of provocation was crossed, the psychological repertoire underwent a qualitative shift, and in response to just a single perturbation step moved from a conciliatory repertoire straight to a sequence of responses characterized by extremely high levels of hostility and open aggression ("hurt him/her as much as possible"). Empirical results show a nonlinear progression of responses from one stable state of positive relations toward another stable state with extremely aggressive relationships. On the other hand, more distant relationships gave rise to more gradual escalation patterns, where mid-range levels of provocation triggered an intermediate responses repertoire. Figure 1 illustrates the general patterns of escalation for the close and distant relationship conditions.

Note that a single static measure at a given moment in time would not predict these paradoxical effects of closeness conditions on the system's dynamics: from a static point of view, close friends are expected to uphold a stable, positive relationship (Hardy, Bukowski, & Sippola, 2002; Cairns, Leung, Buchanan, &

Cairns, 1995). The DST perspective supports this, but only in the case of low levels of provocation, the control parameter in this study.

Additional results from further lines of research show that the trajectory of escalation is critical for later de-escalation: if intermediary states are omitted, and conflict moves from low to high levels of hostility abruptly, then the newly formed attractor for the relationship is less amenable to de-escalation attempts. Results from these studies are a direct validation of the *hysteresis* effect described in the catastrophe theory (Thom, 1994; Poston & Stewart, 1996), and support the catastrophic scenarios of change described by Tesser and Achee (1994). Escalation, from this perspective of the presented approach differs from former static approaches in a number of ways: First, escalation does not only occur at the level of observable behaviors, but also at the level of behaviors that start to be possible within a given interaction, and behaviors that vanish from the range of possibilities. Second, contrary to what the name itself would suggest, conflict escalation does not always occur in a step-by-step fashion, but can be non-linear, and move abruptly from low to high intensity without necessary going through intermediary steps. Third, the "reverse engineering" of escalation processes is not symmetric: the amount of energy needed to "undo" escalating steps is greater than the energy needed to escalate a conflict. Finally, moderate levels of conflict seem to be possible only if no strong pressure is exerted on the system, even if the pressure forces the system into exclusively positive states, so the dynamics of conflict may not stabilize at mid-range levels of intensity.

Further exploration of responses items revealed that dimensions other than destructiveness, as well as triggers for nonlinear dynamics could have emerged as control parameters for the escalation process. For example, trust appeared to be critical in close relationships, and thus in a situation of rupture of trust, a shift has occurred in close relationships conditions, while this factor appeared irrelevant in the distant relationship condition. Such shifts between control parameters are being further investigated, and open an interesting line of research for the study of cultural differences. Further studies using this methodology have been comparing escalation and de-escalation trajectories using individual differences as splitting parameters, and are currently being investigating variance of escalation patterns in various cultural contexts. Additionally, the response scale and the vignettes are currently being adapted in Pakistan, Korea, and Eastern Europe, allowing for the development of more universal behavioral repertoires and provocation stimuli, as well as culture specific sets. Although still in its infancy, a dynamical line of research on changes at the level of behavioral possibilities in response to provocation may shed light on the paradoxical effects of culture on conflict dynamics.

2.3   Culture and Conflict Escalation

The expansion of the DST methodology to the study of cultural factors in conflict dynamics is a work-in-progress. Currently, through in-depth interviews in Egypt,

Iraq, Jordan, Lebanon, Pakistan, Turkey, and the UAE, we are extracting local conflict episodes to develop new stimuli and new behavioral repertoires that are generalizable to the Middle East. Additionally, multidimensional scaling techniques are used to identify main parameters accounting for differences and commonalities in the perception of various conflict behaviors, eventually showing dimensions of disputes across cultures.

Adapted versions of various DST tools, may verify how cultural factors that are relevant to the Middle East, the U.S., and Asia – and in particularly – honor, dignity and face, respectively affect conflict dynamics (Leung & Cohen, 2011; Nisbett & Cohen, 1996). Cultural logics of honor, face, and dignity imply different trajectories of aggression. For example, people in honor cultures have a "keen sensitivity to the experience of humiliation and shame, sensitivity manifested by the desire to be envied by others and the propensity to envy the successes of others" (Miller, 1993, pp. 116). In such cultures, individuals are expected to go to great lengths to uphold the reputation of oneself and one's family and to avoid appearing vulnerable (Vandello & Cohen, 2003). Reputation is critical within cultures of honor, while payback serves as an organizing principle for individuals' interactions when they have been provoked (Delgado, Prieto, & Bond, 1997; Kamir, 2006; Leung & Cohen, 2011; Mosquera, Manstead, & Fisher, 2000; Vandello & Cohen, 2003). Individuals from cultures of honor are thereby expected to respond quickly and with high levels of destructiveness in order to show they are not vulnerable. Individuals are expected to have a quick reaction to even minor acts, due to the importance of reputation maintenance (the premium placed on having a 'tough' reputation necessitates a strong reaction to seemingly small breaches of respect). It is also predicted that individuals will continue to act aggressively towards the perpetrator even after time has passed since the initial transgression; that is, there will be little 'cooling off,' consistent with "hysteresis effects" described in DST (Nowak & Vallacher, 1998). Particular triggers such as damage to female honor, shame, and humiliation are expected to also be important control parameters affecting thresholds of escalation in honor cultures.

In contrast to honor, wherein self-worth can be taken away by other's actions, individuals in dignity cultures are theoretically born with equal worth and rights which cannot be taken away by others (Kamir, 2006; Leung & Cohen, 2011). In dignity cultures, external evaluations matter little, while internal valuations are of the utmost importance. Values such as autonomy, freedom, and standing up for one's beliefs play a crucial role in dignity cultures. Such cultures are also likely to endorse rationality, strong person-task separation, and an independent self-construal (Markus & Kitayama, 1991). We hypothesize that individuals from dignity cultures will generally react to increasingly aggressive acts in a linear fashion. Put differently, the escalation of aggression may be described as a rational, tit-for-tat strategy. In the same manner, individuals should react with decreasing levels of aggression as a perpetrator attempts to de-escalate the situation by apologizing or attempting to restore the relationship. However, particular triggers are expected to result in more severe reactions on the part of a victim from a culture of dignity, such as insults towards one's genuineness and challenges to one's freedom, rights,

and autonomy (Mosquera, Manstead, & Fisher, 2002) or status as an equal member of society (Kamir, 2006), and thus might be seen as critical control parameters of conflict trajectories in dignity cultures.

Finally, in face cultures, individuals place a large premium on external evaluation of the self, while lending little credence to internal evaluations. Face represents an individual's claimed sense of positive image in the context of social interaction (Oetzel & Ting-Toomey, 2003). Similar to honor cultures, upholding the reputation of both the individual and the family is critical. Face cultures tend to have strong norms for communal responsibility, person-task interdependence, and maintaining harmony. Compared to honor and dignity cultures, we expect that individuals from face cultures will be slow to react to initial aggressive acts and will react with less destructiveness. Over time, however, with continued provocation, we expect that conflict dynamics in face cultures can take on a "catastrophic escalation" pattern. Moreover, certain transgressions may trigger strong aggressive reactions in face cultures such as public criticism or embarrassment, communal shame, or violations of duties (Liao & Bond, 2011).

2.4 Discussion

The dynamical-systems approach to the study of culture, negotiation, and collaboration offers the potential to enhance our understanding of the culture and conflict in three distinct ways: metaphorically, mathematically, and empirically. First, dynamical-systems theory offers a rich array of new metaphors, constructs, and principles that might be fruitfully applied to the culture and conflict literature. Dynamic system constructs such as attractors, emergence, and self- organization can serve as useful metaphors to help the researcher understand the dynamic nature of conflict and culture. Second, the dynamical systems approach provides the social scientists tools facilitating the mathematical description of the hypothesized mechanisms underlying specific culture and conflict dynamics. Thus, although social science theory is typically expressed verbally, the dynamical systems tools translate these theories into computer simulations. This will allow identification of assumptions inherent in our theories, but difficult to identify when theories are maintained in their verbal form. Finally, the dynamical-systems approach has implications for the types of empirical methodologies developed and employed in research. Typically, traditional social sciences focus on the central tendency of variables and ignore important dynamics reflected in variables' variances. Further, dynamical-systems models and methods push the social sciences to focus on events as they unfold overtime to understand the general pattern of interactions of the conflict parties overtime.

References

Cairns, R. B., Leung, M. C., Buchanan, L., & Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child development, 66*(5), 1330–1345.

Coleman, P. T., Bui-Wrzosinska, L., Vallacher, R., & Nowak, A. (2006). Approaching protracted conflicts as dynamical systems: Guidelines and methods for intervention. In A. Schneider & C. Honeyman (Eds.), *The negotiator's fieldbook* (pp. 61–74). Chicago: American Bar Association Books.

Coleman, P. T., Vallacher, R., Nowak, A., & Bui-Wrzosinska, L. (2007). Intractable conflict as an attractor: Presenting a dynamical model of conflict, escalation, and intractability. *American Behavioral Scientist, 50*, 1454–1475.

Delgado, A. R., Prieto, G., & Bond, R. A. (1997). The cultural factor in lay perceptions of jealousy as a motive for wife battery. *Journal of Applied Social Psychology, 27*, 1824–1841.

Eckmann, J. P., & Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Review of Modern Physics, 57*(3), 617–656.

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58*, 479–514.

Gottman, J., Swanson, C., & Swanson, K. (2002). A general systems theory of marriage: Nonlinear difference equation modeling of marital interaction. *Personality and Social Psychology Review, 6*, 326–340.

Granic, I., & Lamey, A. V. (2002). Combining dynamic systems and multivariate analyses to compare the mother-child interactions of externalizing subtypes. *Journal of Abnormal Child Psychology, 30*, 265–283.

Guastello, S., Koopmans, M., & Pincus, D. (Eds.). (2008). *Chaos and complexity in psychology: The theory of nonlinear dynamical systems*. Boston: Cambridge University Press.

Hardy, C. L., Bukowski, W. M., & Sippola, L. K. (2002). Stability and change in peer relationships during the transition to middle-level school. *The Journal of Early Adolescence, 22*(2), 117–142.

Imai, L., & Gelfand, M. J. (2010). The culturally intelligent negotiator: The impact of cultural intelligence (CQ) on negotiation sequences and outcomes. *Organizational Behavior and Human Decision Processes, 112*, 83–98.

Kamir, O. (2006). Honor and dignity in the film forgiven: Implications for sociolegal theory. *Law & Society Review, 40*, 193–234.

Leung, A. K.-Y., & Cohen, D. (2011). Within- and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of Personality and Social Psychology, 100*, 507–526.

Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45*, 587–597.

Lewis, M. D. (2000). The promise of dynamic systems approaches for an integrated account of human development. *Child Development, 77*, 36–43.

Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science, 2*, 457–475.

Liao, Y., & Bond, M. H. (2011). The dynamics of face loss following interpersonal harm. *Journal of Cross-Cultural Psychology, 42*, 25–38.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253.

Miller, W. I. (1993). *Humiliation: And other essays on honor, social discomfort, and violence*. Ithaca, NY: Cornell University Press.

Mosquera, P. M. R., Manstead, A. S. R., & Fischer, A. H. (2000). The role of honor-related values in the elicitation, experience, and communication of pride, shame, and anger: Spain and the Netherlands compared. *Personality and Social Psychology Bulletin, 26*, 833–844.

Mosquera, P. R., Manstead, A., & Fischer, A. H. (2002). The role of honour concerns in emotional reactions to offences. *Cognition and Emotion, 16*, 143–163.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Denver, CO: Westview Press.

Nowak, A., & Lewenstein, M. (1994). Dynamical systems: A tool for social psychology? In R. R. Vallacher & A. Nowak (Eds.), *Dynamical systems in social psychology* (pp. 17–53). San Diego, CA: Academic.

Nowak, A., & Vallacher, R. R. (1998). *Dynamical social psychology*. New York: Guilford Press.

Oetzel, J. G., & Ting-Toomey, S. (2003). Face concerns in interpersonal conflict: A cross-cultural empirical test of the face negotiation theory. *Communication Research, 30*, 599–624.

Poston, T., & Stewart, I. (1996). *Catastrophe theory and its applications* (Vol. 2). Dover Publications.

Schuster, H. G. (1984). *"Deterministic chaos" an introduction* (Monographie 248 Seiten). Weinheim, Germany: Physik.

Tesser, A., & Achee, J. (1994). Aggression, love, conformity and other social psychological catastrophes. In R. Vallacher & A. Nowak (Eds.), *Dynamical systems in social psychology* (pp. 95–109). San Diego, CA: Academic.

Thom, R. (1975). *Structural stability and morphogenesis. An outline of a general theory of models*. Reading, MA: W. A. Benjamin, Inc.

Thom, R. (1994). *Structural stability and morphogenesis*. Westview Press.

Vallacher, R. R., Coleman, P. T., Nowak, A., & Bui-Wrzosinska, L. (2010). Rethinking intractable conflict: The perspective of dynamical systems. *American Psychologist, 65*(4), 262–278.

Vallacher, R. R., & Nowak, A. (2007). Dynamical social psychology: Finding order in the flow of human experience. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 734–758). New York: Guilford Publications.

Vallacher, R., Nowak, A., Bui-Wrzosinska, L., & Coleman, P. T. (2006). Attracted to conflict: A dynamical perspective on malignant social relations. In A. Golec & K. Skarzynska (Eds.), *Understanding social change: Political psychology in Poland* (pp. 33–50). Haauppague, NY: Nova Science Publishers Ltd.

Vandello, J. A., & Cohen, D. (2003). Male honor and female fidelity: Implicit cultural scripts that perpetuate domestic violence. *Journal of Personality and Social Psychology, 84*, 997–1010.

A Game-Theoretic Approach to Modeling Cross-Cultural Negotiation

Miroslav Dudík and Geoffrey J. Gordon

Abstract Faithful models of negotiation should capture aspects such as subjective incentives, imperfect information, and sequential interaction, while providing explanation for behaviors such as bluffing, trust building, and information revelation. All of these objectives are elegantly addressed by theory of sequential games, and some of these phenomena have no convincing explanation without game theory's key assumption, namely, that of the rationality (or approximate rationality) of the negotiators. In this paper we discuss a game-theoretic approach to modeling negotiation. In addition to accounting for a range of behavior and reasoning styles we also address several aspects specific to cross-cultural negotiation. We argue that the existence of culture-specific beliefs and strategies can be explained by the existence of multiple game-theoretic equilibria. Within a culture, repeated interaction and learning lead to an equilibrium. On the other hand, across cultures, infrequent interaction leads with high probability to disparate (and often incompatible) equilibria. We hypothesize that inefficiency in cross-cultural negotiation can be attributed to this incompatibility. We discuss recently-developed algorithms that can be used to fit models of culture-specific behavior from data while incorporating rationality constraints. We anticipate that the additional structure imposed by rationality constraints will yield both statistical advantages and game theoretic insights.

Keywords Game theory • Communication • Negotiation • Models of cultural differences • Multi-agent influence diagrams (MAIDs) • Extensive form correlated equilibrium

M. Dudík (✉)
Microsoft Research, 1290 Avenue of the Americas, 6th Floor, NY 10104, New York
e-mail: mdudik@microsoft.com

G.J. Gordon
Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue,
Pittsburgh, PA, 15213, USA
e-mail: ggordon@cs.cmu.edu

1 Introduction

Negotiation among multiple parties with the goal to establish and maintain business relationships or resolve political conflicts is an important problem studied both by economists and behavioral scientists (Ausubel, Cramton, & Deneckere, 2002; Lax & Sebenius, 1986; Pruitt & Lewis, 1975; Rubinstein, 1982; Thompson, 1990; Walton & McKersie, 1965). Computational modeling efforts typically either focus on statistical modeling of the opponent (Ficici & Pfeffer, 2008; Ray, King-Casas, Montague, & Dayan, 2009), or posit a specific mode of interaction (protocol) according to which negotiating parties exchange information (Chalkiadakis & Boutilier, 2007; Lai, Sycara, & Li, 2008). Both of these computational approaches fail to fully account for the fact that opponents are aware that we are optimizing our own objectives, and aim to use this information strategically to optimize their own objectives. This aspect of negotiation is addressed by concepts of game-theoretic equilibria. In this paper we outline the challenges of modeling negotiation and show how game-theoretic treatment addresses them. We also show how game theory accounts for the rise of culture-dependent beliefs about others, and explains the inefficiency in cross-cultural negotiation.

By modeling a negotiation, we mean the effort to capture a wide range of observed *behaviors* as well as *reasoning styles* of individual agents—behaviors include actions such as "sharing information" (e.g., communicating one's preferences), whereas reasoning includes internal judgments such as "the other agent just offered a compromise, so she probably values fairness". Since both behavior and reasoning take place in concrete situations that arise over the course of multi-agent interaction, a successful modeling approach needs to account for the structure of this interaction. Two key aspects of the interaction are the *imperfect information*—for example agents initially know only their own objectives, but have only probabilistic expectations about the objectives of others—and *sequential decisions*, for example, speaking turns.

A partial solution to the previous challenges is provided by *partially observable Markov decision processes* (POMDPs) and similar latent-state sequential models (Ficici & Pfeffer, 2008; Ray et al., 2009). They take a view of a single agent who is trying to optimize her own utility function in a sequential interaction with imperfectly observed environment which includes all other agents. The key limitation of POMDPs is that they do not capture *strategic reasoning*. Strategic reasoning accounts for the fact that the environment includes other agents who are trying to optimize their own utility functions. While POMDPs can contain latent state variables describing opponents' beliefs and policies (an approach known as *opponent modeling*), a POMDP cannot account for strategic behavior such as randomization among moves. Randomization is essential to human-like behavior: e.g., a human would never play a deterministic strategy in rock-paper-scissors. In order to give a principled account of strategic behavior, we will therefore model negotiation as an extensive-form game. Extensive-form games capture both sequential structure of interaction and imperfect information while providing a wide range of behaviors and reasoning that arise solely as a result of the rationality of an agent.

Fig. 1 An example game-theoretic model of negotiation between two agents
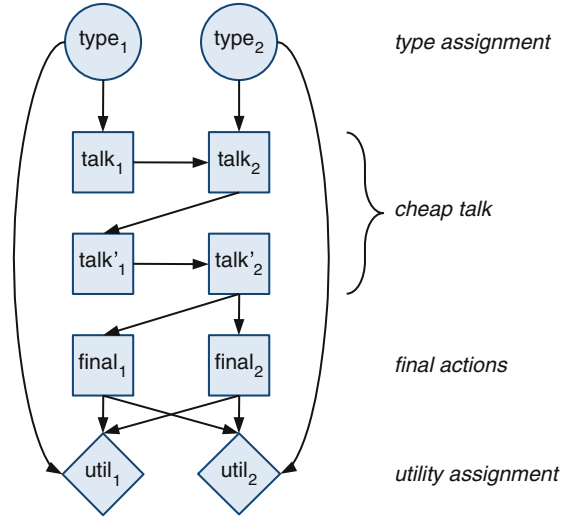
In Sect. 2 we describe extensive-form games in more detail. However, in practice we usually work with their compact representations such as multi-agent influence diagrams (MAIDs), also described in Sect. 2. In Fig. 1, we give an example of a MAID model of negotiation. Each agent has a privately known objective, formally denoted as the type, which may include aspects such as collectivism or individualism. Knowing their types (but not the types of others), agents take turns talking to each other and at the end they make their final decision (take the final action). The payoff (or utility) that they derive from the outcome depends on their type, their final action and on the final action of others. Arcs in the graph denote these dependencies—for example, when the second agent speaks, she is aware of her type as well as of what the first agent just said. Some additional arcs are implied by the condition of the perfect recall (agents do not forget anything throughout the game)—for example, when the second agent speaks for the second time, she continues to be aware of her type and all that was said before. The purpose of the communication is to foreshadow the final action and achieve some coordination.

Besides modeling the process of negotiation, we are particularly interested in modeling negotiation among participants from different cultures. In particular, we seek explanations for the inefficiency of cross-cultural negotiation. Using our game-theoretic setup, we hypothesize that this inefficiency can be attributed to the existence of multiple equilibria in negotiation games. One of the key problems in maintaining an equilibrium is the problem of believable and reliable information sharing. There are many possible ways for a culture to solve the information sharing problem. Within a culture, repeated interaction and learning lead to an equilibrium. On the other hand, infrequent interaction across cultures means that different cultures will likely converge to different (and often incompatible) equilibria.

While game theory addresses issues that cannot be addressed by opponent modeling, its assumptions are somewhat idealized. Specifically, perfect rationality and perfect recall are not accurate descriptions of real players. Yet, we believe that by capturing salient aspects of behavior game theory offers highly suitable models for analyzing negotiation. In the remainder of this paper we outline our modeling approach in more detail and discuss how game-theoretic equilibria yield various reasoning and behavior styles.

2   Game-Theoretic Setup

We study a sequential game-theoretic formalism known as extensive-form games (EFGs) (for details, see, e.g., Fudenberg & Tirole, 1991). Let N be the number of agents, denoted as $n = 1, \ldots, N$. An extensive-form game is represented by a *game tree*, where non-leaf nodes represent either randomness (an action by nature) or an action by one of the agents. Non-leaf nodes of agents are partitioned into *information sets*. In each information set i a unique agent n is required to act. The agent knows only the identity of the information set, but cannot distinguish among the nodes in it, which represents partial information. Nodes in the same information set have the same set of available actions. For each such action a, there is an edge labeled a leaving every node in the information set. The game begins in the root of the tree; agents and nature take turns until reaching a leaf. Each leaf contains an assignment of utilities to individual agents. The collection of information sets of an agent n is denoted $I(n)$. A deterministic behavior of the agent n is described by a *pure strategy*, which is a tuple $s_n = (s_i)_{i \in I(n)}$ with s_i specifying which action to take in each information set. The vector $s = (s_n)_{n \leq N}$ of pure strategies of all agents is referred to as the *strategy profile*. Nature's actions are drawn from some predetermined distributions.

As common in game theory, we restrict our attention to EFGs with *perfect recall*. Perfect recall means that agents do not forget any information over the course of the game. Formally, this requires that paths reaching nodes in an information set $i \in I(n)$ are indistinguishable by n, i.e., they contain identical sequences of n's information sets, and the agent n took identical actions in those information sets.

Instead of working directly with the game-tree representation, we work with a more succinct representation called *multi-agent influence diagrams* (MAIDs) (Koller & Milch, 2001). Multiagent influence diagrams are game-theoretic generalizations of Bayes nets. Similarly to Bayes nets, MAIDs are represented by directed acyclic graphs. They have three types of nodes: (i) *decision nodes* (represented as rectangles), where a specified agent assigns a variable given the values of the parent variables, (ii) *chance nodes* (represented as ovals), where nature randomly assigns a variable conditioned on the values of the parent variables (according to a specified conditional distribution), (iii) *utility nodes* (represented as diamonds), where a specified agent receives utility as a function of the values of the parent variables; utility nodes have no outgoing edges. Any topological order specifies the

order in which the play proceeds, for each node, distinct assignments of its parent variables correspond to distinct information sets.

To model negotiation, we consider MAIDs with a structure similar to the one in Fig. 1. First, nature assigns to each agent a *type* representing their utility function. As the arcs indicate, agents know their own types but not the types of others. Type assignments are followed by several rounds of *cheap talk*: "talk", because actions are fully observed, and "cheap", because actions do not directly result in any utility. After cheap talk, each agent carries out the *final action*. The combination of final actions of all agents determines the utility each agent receives.

Some strategic behavior, such as trust building, needs to be modeled in the context of a repeated interaction. In that case, the final action is followed by each agent's noisy observation of the final actions of other agents, after which more rounds of cheap talk ensue followed by another final action, etc. We assume that the type assignment does not change between the stages.

The game-theoretic structure outlined so far captures only the structure of the interaction (sequential decisions and imperfect information). To capture the reasoning and especially strategic reasoning, we next shift our attention to equilibria in extensive-form games.

3   Equilibrium Reasoning

The most common equilibrium concept is that of Nash equilibrium. Unfortunately, many questions related to finding Nash equilibria are intractable, which we believe casts doubt on their suitability as behavioral models. Instead we will work with *extensive-form correlated equilibria* (EFCEs) (von Stengel & Forges, 2008), which can be computed more efficiently and, in addition, can be achieved as results of repeated interaction among learning agents (Gordon, Greenwald, & Marks, 2008). This latter property lends support to the claim that EFCEs are a suitable model of negotiation behavior among same-culture agents. (We will return to cross-cultural negotiation below.)

Formally, an EFCE is a probability distribution over strategy profiles implemented by a moderator as follows. As information sets are reached during the game, the moderator suggests an action privately to the corresponding agents. Agents can either follow the moderator's suggestion or deviate. When agents deviate, they stop receiving suggestions from the moderator and must follow their own strategies. The probability distribution $p(s)$ is an EFCE if none of the agents can benefit by deviating.

The equilibrium distribution $p(s)$ determines the equilibrium reasoning of agents. For example, if the game reaches the information set i, belonging to the agent n, and the moderator suggests to take the action a then the agent n can calculate the posterior distribution $p(s_{-n} \mid s_i = a)$ which corresponds to the belief about other agents. Under this belief, the agent cannot improve her payoff by deviating from the suggestion a. This type of behavior will indeed be

optimal, conditioned on the other agents' strategies, provided that the equilibrium distribution is the common knowledge (and thus beliefs are mutually compatible).

4 Cross-Cultural Negotiation

In general, there are many EFCEs, and because of limited cross-cultural interaction we hypothesize that different cultures converge to different EFCEs. Using the algorithm of Dudík and Gordon (2009), it is possible to find EFCEs consistent with given negotiation transcripts. Negotiation transcripts will typically contain some information about agent types (agent personalities or details of their role), cheap talk (coded speaking turns), and final actions (the outcome of the negotiation). Given a set of same-culture negotiation transcripts, we can try to fit an equilibrium that matches the observed frequencies of various events such as "providing information by collectivist agents at the beginning of the game", or "proposing a multi-issue offer within two rounds after another agent describes their priorities". If our hypothesis is correct, the inclusion of equilibrium constraints should yield more accurate estimates of reasoning patterns (measured as beliefs at different points in the game) compared with the predictions of other latent-state models.

The analysis of what happens in cross-cultural negotiation is more open. Because of infrequent contacts, it is unlikely that cross-cultural interactions are at an equilibrium. In a somewhat artificial case, when the agents are not aware that the other parties come from different cultures, they may continue to apply the strategy learnt for their own culture. Since such strategies are unlikely to be in an equilibrium across cultures, the agents are open for exploitation.

Even when agents can identify that others belong to a different culture, their strategy may rely on patterns learnt in the same-culture negotiation. These are modified by (i) their previous contact with the other culture, and (ii) their prejudice. The question how these factors combine, as well as to what extent the mutual prejudices are in an equilibrium is open for future research.

References

Ausubel, L., Cramton, P., & Deneckere, R. (2002). Bargaining with incomplete information. In R. Aumann, & S. Hart (Eds.). *Handbook of game theory* (Vol. 3, pp. 1897–1945). Amsterdam: Elsevier.

Chalkiadakis, G., & Boutilier, C. (2007). Coalitional bargaining with agent type uncertainty. In *Proceedings of the 20th international joint conference on artificial intelligence*. Hyderabad: India.

Dudík, M., & Gordon, G. J. (2009). A sampling-based approach to computing equilibria in succinct extensive-form games. In *Proceedings of the 25th annual conference on uncertainty in artificial intelligence*. Montreal, Quebec: Canada.

Ficici, S., & Pfeffer, A. (2008). Modeling how humans reason about others with partial information. In *Proceedings of the 7th international conference on autonomous agents and multi-agent systems*. Estoril: Portugal.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.

Gordon, G. J., Greenwald, A., & Marks, C. (2008). No-regret learning in convex games. In *Proceedings of the 25th international conference on machine learning*. Helsinki: Finland.

Koller, D., & Milch, B. (2001). Multi-agent influence diagrams for representing and solving games. In *Proceedings of the 17th international joint conference on artificial intelligence*. Seattle: Washington, USA.

Lai, G., Sycara, K., & Li, C. (2008). A decentralized model for automated multi-attribute negotiations with incomplete information and general utility functions. *Multiagent and Grid Systems, 4*(1), 45–65.

Lax, D. A., & Sebenius, J. K. (1986). *The manager as negotiator: Bargaining for cooperation and competitive gain*. New York: Free Press.

Pruitt, D. G., & Lewis, S. A. (1975). Development of integrative solutions in bilateral negotiation. *Journal of Personality and Social Psychology, 31*(4), 621–633.

Ray, D., King-Casas, B., Montague, P. R., & Dayan, P. (2009). Bayesian model of behaviour in economic games. In *Advances in neural information processing systems 21*. Vancouver, BC: Canada.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica, 50*(1), 97–109.

Thompson, L. L. (1990). An examination of naive and experienced negotiators. *Journal of Personality and Social Psychology, 59*(1), 82–90.

von Stengel, B., & Forges, F. (2008). Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research, 33*(4), 1002–1022.

Walton, R. E., & McKersie, R. B. (1965). *A behavioral theory of labor negotiation: An analysis of a social interaction system*. New York: McGraw-Hill.

Inter-cultural Opponent Behavior Modeling in a POMDP based Automated Negotiating Agent

Praveen Paruchuri, Nilanjan Chakraborty, Geoff Gordon, Katia Sycara, Jeanne Brett, and Wendi Adair

Abstract As the world gets increasingly networked, business and political negotiations take place between people of different cultures. Cross-cultural negotiations have been mainly studied empirically and there is a dearth of computational models of negotiation that incorporate the culture of the negotiators. In this chapter, we take the first steps towards building a Partially Observable Markov Decision Process (POMDP) based Automated Negotiation (PAN) agent, that takes the culture of the negotiators into account. We move away from the offer-counteroffer paradigm that is usually used in computational modeling of negotiation. We assume that apart from making offers, the agents can take other actions for seeking/providing information during negotiation. Our approach is motivated by the experimental findings that (a) during negotiation, people communicate their preferences and justification of their preferences apart from making direct offers and (b) cultural distinctions can be made between negotiating agents based on an abstract coding of their conversation. We show that in accordance with an existing cognitive theory of inter-cultural negotiation from behavioral psychology literature, we can construct a POMDP model of negotiation. A key challenge in developing the PAN agent is in obtaining the state transition function for the POMDP. We demonstrate that the state transition function can be built from transcripts of actual negotiations between people.

P. Paruchuri (✉) • N. Chakraborty • G. Gordon • K. Sycara
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA
e-mail: paruchur@gmail.com; nilanjan@cs.cmu.edu; ggordon@cs.cmu.edu; katia@cs.cmu.edu

J. Brett
Kellogg School of Management, North Western University, Evanston, IL 60208, USA
e-mail: jmbrett@kellogg.northwestern.edu

W. Adair
Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada
e-mail: wladair@uwaterloo.ca

1 Introduction

Negotiation is an interactive process by which multiple parties with limited common knowledge try to arrive at an agreement over a set of issues with possibly conflicting preferences over the issues. The topic of negotiation has received wide attention across various fields such as political science (Druckman, 1986), economics (Myerson, 1977), behavioral sciences (Pruitt, 1981), computational sciences (Kraus, 1997) etc. For example, political scientists are interested in analyzing and predicting the negotiation processes between various countries and cultures while computer scientists are more interested in building computational models that can be embodied in automated agents. Most work in computational modeling to-date use Rubinstein's alternating offer model for bargaining games (Rubinstein, 1982) and focuses on the outcome of a negotiation. The cultural background of the players involved in the negotiation is usually ignored. However, empirical studies (see Gelfand & Brett, 2004 and references therein) with human negotiators have established that culture plays a crucial role in the way a negotiation progresses and on the outcome of negotiation. In this chapter, we propose a computational modeling approach for a general purpose negotiation problem that takes the culture of the players into account.

Although the offer-counteroffer paradigm is a popular way for computational modeling of negotiations, empirical studies indicate that during negotiation, apart from making direct offers, people use other forms of information exchange. For example, people provide qualitative information about their preferences, defend their positions, show positive or negative reactions to their opponents comments etc. Moreover, cultural attributes of the players influence the way preferences are expressed and interpreted. Therefore, in this study, we move away from the offer-counteroffer paradigm and assume that the negotiators can take other actions for seeking/providing information during negotiation. We assume that the players take turn alternately in taking their actions (i.e., making offers or seeking/providing/justifying preferences). Since the utterances of people during negotiation can be in any natural language and natural language interpretation is not a mature technology yet, we abstract the utterances as belonging to a set of codes. These codes encode the behavior used by the negotiator, like asking questions, providing information, making offers, defending positions etc. These codes are used by researchers in the behavioral psychology literature for studying negotiation. Each utterance can correspond to one or more codes and these codes form the set of actions that can be taken by each negotiating agent.

We model the negotiation problem in a decision-theoretic framework using the Partially Observable Markov Decision Process (POMDP) (Kaelbling, Littman, & Cassandra, 1998). The advantages of a POMDP-based modeling approach are as
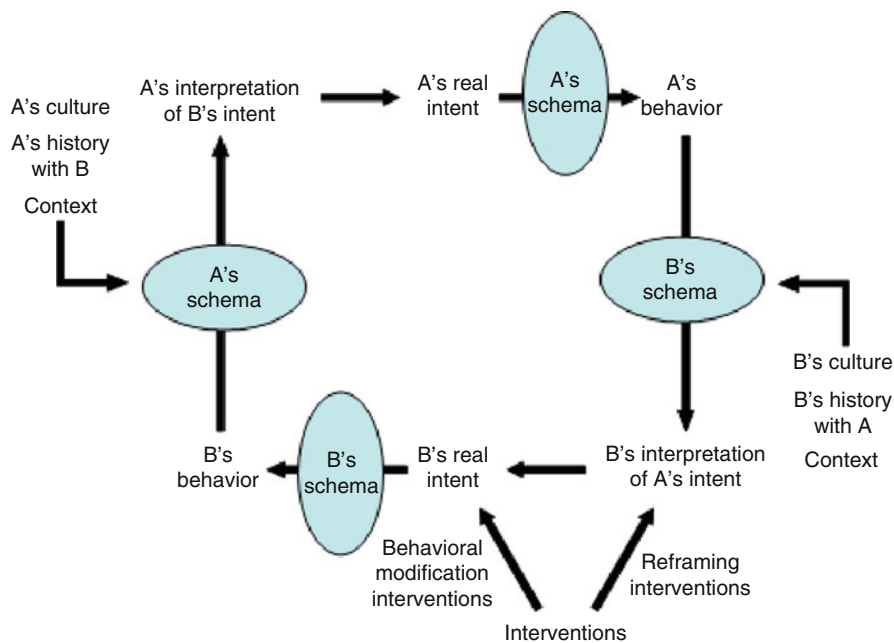
Fig. 1 Cognitive schema in dynamic collaboration and negotiation

follows: (a) A POMDP based approach is decentralized, i.e., each agent solves her own POMDP model while maintaining a belief about the other agents. (b) POMDPs provide a natural way to capture the sequential nature of the negotiation process while incorporating the new observed data such as the other agent's actions; additionally POMDPs provide ways to refine an agent's belief about other agents. (c) POMDPs can incorporate the effect of cultural factors in a natural way, e.g., given the action of the opponents and their cultural background, it uses the knowledge in interpreting the actions of the opponent, or in deciding the best action to take (please see the discussion in the next paragraph). A key challenge in developing the POMDP based Automated Negotiation agent (PAN agent) is in designing the state transition function of the POMDP. We therefore focus in this chapter on describing the various steps involved in building a realistic transition function from transcripts of actual negotiation between people. We use logistic regression to compute the state transition function and present experimental results to demonstrate that the computed transition function can distinguish between different cultures, given the transcript of the negotiations.

Figure 1 illustrates a common way of representing how cognitive schemata (Tinsley, 2001) change during negotiations. The figure shows how schema's filter and interpret incoming stimuli and guide outgoing reactions for a simple two-party (say, *A* and *B*) interaction, e.g., for resolving a conflict. Party *A*'s schema enters in two places. First, it is the lens through which party *B*'s behavior will be interpreted,
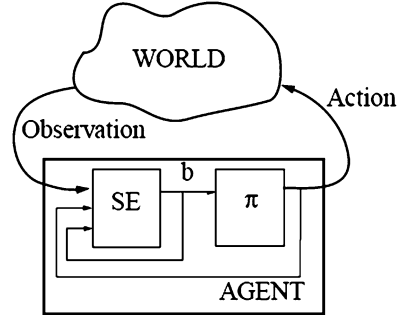
and second it is the filter through which A's actual intentions will give rise to concrete behaviors visible to the other party. In this way, schema's become relevant whenever an individual is taking information from the outside world or offering behavior to the outside world. A's culture and history of interaction with B (or members of B's culture) will influence party A's schema. Important components of a schema are goals (what is appropriate to try to achieve), norms (what is appropriate behavior to go about getting what you want), and beliefs and attributions about the character of the other person. A's schema includes "who B is", which influences A's interpretation of B's behavior of "what B is doing". This drives A's intentions or strategies for subsequent moves (e.g., should A be cooperative or not). A's intentions will then drive A's behavior, as filtered again through A's schema, which, includes norms for appropriate behavior. Since this is a symmetric situation, B's schema filters A's behavior and influences how B interprets that behavior, which influences how B intends to respond.

We now provide a brief intuition on how the cognitive schema presented in Fig. 1 naturally fits into a POMDP model. To illustrate this, we provide here an informal description of the POMDP model while a more formal description is provided in Sect. 3. Figure 2 (taken from Kaelbling et al., 1998) shows the working of a POMDP. The world of the POMDP is composed of states. Initially, an agent believes that it is in a particular state or has a probability distribution over states, called the belief state. The agent takes an action and gets an observation of the new state she reaches. Given that the new state may not be directly inferred from the observation, the state estimator (labeled SE) derives the new belief state based on the last action, the current observation and the previous belief state. Once the new belief state is calculated, the agent takes a new action and the process continues till an end state is reached. The block labeled π represents the POMDP policy. The policy of the POMDP maps a belief state to an action. We now model the POMDP from the perspective of party A for the cognitive schema presented in Fig. 1. The context consisting of A's culture and A's history with B contribute towards modeling the initial belief of the POMDP model. A's observation of B's actual behavior maps to the observation set, while A's schema and real intent which are part of A's belief define a belief distribution over the state space. A's behavior

gets captured in the action set of the POMDP. The perceived model of B (learnt using regression techniques described later) becomes the transition function of A's POMDP. Similar mapping can be done for agent B. This direct mapping between a general purpose cognitive schema for negotiation and a POMDP model, reinforces our belief that modeling the negotiation problem as a POMDP may be a good approach to follow.

2 A General Purpose Negotiation Setup

We now introduce the general purpose negotiation setup that we consider in this paper. In Sect. 3.1, we will present a simple example to illustrate the POMDP based model discussed below. Consider $i \in \{1, 2, .., n\}$ self-interested agents negotiating a set of issues $j \in \{1, 2, .., m\}$. Given that each agent is inherently different, they have different expectations on what a fair solution is, what a fair way to negotiate would be etc. The differences between each agent can be captured through the notion of types. We assume that each agent has a type $\tau \in \{\tau_1, \tau_2, \ldots, \tau_u\}$ based on which she acts. Furthermore, we assume that an agent, i, knows her own type τ_i while it has a probability distribution over the opponent types. The type of an agent is a function of the various social and cultural factors that are used in behavioral psychology to model the socio-cultural background and social value orientation of an individual (Camerer, 2003).

Some of the cultural factors identified in the empirical literature on human negotiation (see Brett, 2000; Drake, 2001) include individualism versus collectivism, egalitarianism versus hierarchy, direct versus indirect communication and distributive (D) versus integrative (I) approach. Different cultures attach different degrees of importance to these factors and that affects the negotiation strategies people use. For example, Japanese culture gives more importance to collectivism whereas American culture is more individualistic. Social value orientations (i.e., goals that individuals may pursue for making decisions when the outcome may affect others) like altruism, cooperativeness, individualism, competitiveness, and aggressiveness have also been identified as factors that affect negotiation (Camerer, 2003). For purposes of this chapter, we will consider only the cultural factors. We do not go into the specifics of each cultural factor but keep the modeling at a higher level, i.e., whether the negotiator is Japanese, or American, or Israeli. Our goal is to build an automated negotiating agent that can adjust its belief regarding the opponent's culture and choose appropriate actions to reach a high quality negotiated agreement. The assumption here is that initially the agent has little idea of the culture of the other party and hence we model it as a uniform initial belief. Over the course of the negotiation the agent may realize that the initial belief it held about the other agent's type may not be true and refines the belief as the negotiation progresses. We now describe how the POMDP framework captures the general negotiation process just presented.

3 The Partially Observable Markov Decision Process Framework

In Sect. 1 we provided an informal description of the POMDP model. Formally, a POMDP can be defined by the tuple $\{S, A, T, \Omega, O, R\}$, where S is a finite set of states; A is a finite set of actions; $T(s, a, s')$ provides the probability of transitioning from state s to s' when taking action a; Ω is a finite set of observations; $O(s', a, o)$ is the probability of observing o after taking an action a and reaching state s' and $R(s, a)$ represents the reward function, i.e., the reward obtained by taking action, a, at state, s. While modeling the POMDP we also need to specify an initial belief where a belief state b is defined as a probability distribution over the set of states S. Once the negotiation problem is cast into the POMDP framework, many algorithms both heuristic and exact exist in the literature to find the approximate/optimal POMDP policies (Kaelbling et al., 1998; Pineau, Gordon, & Thrun, 2003). A policy here refers to the mapping between a belief state (which includes the observation at that time-step) to an action. Effectively, the agent solves the POMDP and obtains a policy table. The agent can then use this table to decide on the appropriate action to take at each time-step based on the state the agent is in.

The main challenge in casting a negotiation problem as a POMDP lies in defining the tuple $\{S, A, T, \Omega, O, R\}$ for an instance of the problem. The state space S can be defined using the knowledge of the problem domain and the various factors affecting negotiation that have been identified in the behavioral sciences literature (Pruitt, 1981). The action space A and the set of observations Ω can also be formed using domain knowledge as well as knowledge about the strategies used by people in a negotiation (that has been identified in the behavioral sciences domain). From the point of view of modeling, the main hurdle lies in coming up with the appropriate parameters for the state transition function, T, observation function O, and the reward function R. In the next section we describe a simple negotiation problem and the procedure to build the POMDP with a specific focus on the transition function.

3.1 Encoding a General Negotiation Problem in the POMDP Framework

We now present a simple negotiation example based on the Cartoon data (Adair & Brett, 2005) to illustrate various aspects of the POMDP model. The Cartoon data is multi-cultural version of the Mom.com (Tenbrunsel & Bazerman, 1995) exercise about the sale of syndication rights for 100 episodes of children's cartoon named Ultra Rangers. For purposes of encoding using a POMDP, we abstract away the details of this data and present a simplified version here. The actual version of this data is used in our experiments section. In the simplified example, we have two agents namely $A1$ and $A2$, bargaining over one issue. Here $A1$ is the seller

and $A2$ the buyer. We assume that the best value for the seller is 100 units while it is 0 units for the buyer. Note that this scale captures a general set of scenarios since any other scale can be normalized and shifted to fit in this. The corresponding worst case scenario is 0 for $A1$ and 100 for $A2$. For purposes of this analysis, as mentioned earlier we assume that a single factor namely the culture determines the type of all the agents. We model this factor as a binary random variable, 0 and 1 for two different cultures. Since the POMDP agent cannot observe this factor directly, she will construct a belief about it, represented as a probability between 0 and 1. In the following subsection, we describe the procedure to construct the POMDP for $A1$.

3.2 State Space Definition

S is the set of world states. For the negotiation problem, the state vector of the POMDP has the following factors:

My Type, Opponent Type, My last proposal, Opponent last proposal, Previous n observations.

Formally, $A1$'s state at time step t can be represented as:

$$s_1^t = \langle \tau_1, \tau_2, p_1^{t-1}, p_2^{t-1}, \langle a_2^{t-1}, a_2^{t-2}, \ldots, a_2^{t-n} \rangle \rangle$$

where,

- τ_1 is agent $A1$'s type
- τ_2 is agent $A2$'s type
- p_1^{t-1} is the most recent $A1$'s proposal until time step $t-1$ (Note that proposal values of 100 and 0 are assumed to be the on the table for $A1$ and $A2$ respectively at $t = 0$)
- p_2^{t-1} is the most recent $A2$'s proposal till time step $t-1$ and
- a_2^{t-n} is $A2$'s action at time step $t-n$.

The problem with the above state definition is that the transition from a state s_1^t to a state s_1^{t+1} is determined by the observation a_2^t, thus violating the Markov property. To make the transition to state s_1^{t+1} independent of the observation, we include a single future observation in the state s_1^t. A POMDP that includes a single future observation in the state i.e., n = 1, is called a *first* order POMDP. For an arbitrary value of n, we call it a *n*th order POMDP. Hence the new state vector can be represented as:

$$s_1^t = \langle \tau_1, \tau_2, p_1^{t-1}, p_2^{t-1}, a_2^t, \langle a_2^{t-1}, a_2^{t-2}, \ldots, a_2^{t-n} \rangle \rangle$$

However this future observation at each state is unknown and needs to be predicted (procedure to do this is described later). For purposes of keeping the state space

of the POMDP small, we model only the opponent type which we assume plays a much stronger role than the agent type itself. Hence the new state vector would be:

$$s_1^t = \langle \tau_2, p_1^{t-1}, p_2^{t-1}, a_2^t, a_2^{t-1} \rangle$$

3.3 Action Set and Observation Set

The set of actions in our domain can be classified as information actions and proposal actions. Proposal actions capture the value of issue being discussed while information actions represent the general chat that happens between the two players. A detailed coding scheme was developed for purposes of the Cartoon data containing 34 codes. To limit the size of the POMDP we generate, we use a higher level coding scheme containing 11 codes corresponding to the following actions:

Information, Questions, Substantiation, Single-issue offer, Multi-issue offer, Multi-issue offer with trade-off, Reactions, Mutuality, Procedural comments, Miscellaneous, Confirmation.

The set of observations Ω for the agent corresponds to the action set of the buyer. For our purposes, we assume that both the players have the same action set described above and therefore the observation set for both agents is also the same.

3.4 State Transition Function and Observation Function

The transition function represents the probability with which an action a taken from a state s^t leads to a set of states s^{t+1}. Transitions in our domain are stochastic. Let the state s for agent $A1$ at time t be:

$$s_1^t = \langle \tau_2, p_1^{t-1}, p_2^{t-1}, a_2^t, a_2^{t-1} \rangle$$

Let the action at state s_1^t be a_1^t. The agent 2's type (factor 1 of the state vector) τ_2 in s^{t+1} remains the same as in s^t. This is because the ground truth about agent 2's type remains the same, i.e., if $A2$ is of type 0 she/he will remain of type 0. If action a_1^t is a proposal action then the factor p_1^{t-1} changes in the new state and remains same otherwise. Similar reasoning holds for p_2^{t-1} where the value changes only if $A2$'s action a_2^t is a proposal. The stochastic part of the transition function is due to the factor a_2^t, where $A1$ needs to predict the future action of $A2$, a_2^{t+1} which is the observation for $A1$ at the state s_1^{t+1}. The final factor of the state is the previous action of $A2$, a_2^{t-1} which $A1$ obtains as the observation at state s_1^{t-1}. Figure 3 shows a sample POMDP state space along with possible transitions.

Given that the state captures the previous observation, the observation function is deterministic and is mathematically represented as follows:
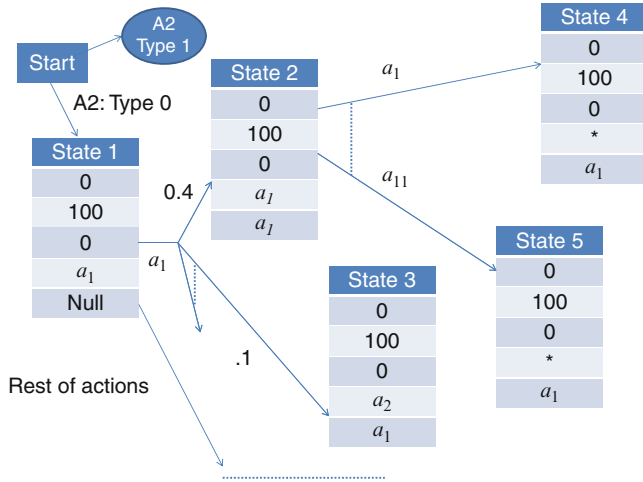
Fig. 3 Partial graph for a simple negotiation example encoded as a POMDP

$$pr(o^t|s_1^{t+1}, a_1^t) = pr(o^t|s_1^{t+1}) = pr(o^t|a_2^t)$$
$$= 1, if o^t = a_2^t \text{(captured as previous observation in } s_1^{t+1})$$
$$= 0, \text{otherwise}$$

3.5 Reward Function: R

Given our emphasis on building a realistic transition function, we do not present the reward function here. The rest of the components were needed to perform the belief tracking experiments presented in the next section and hence modeled here.

4 Computing the State Transition Function

The transition function just described is learnt using a logistic regression with regularization (Hilbe, 2009). Logistic regression is a mathematical technique that is used for prediction of the probability of occurrence of an event by fitting data to a logit function. We are using the WEKA machine learning package (Witten & Frank, 2005) for purposes of experimentation. The input to the logistic regression function are three matrices X, Y and X′. The X matrix is a n by p matrix of p predictors at each of the n observations. Y is a n by 1 matrix of one outcome for each of the n observations. X′ is n′ by p matrix of p predictors for each of n′ observations where n′ is obtained as a Cartesian product of the maximum value of each of the p

predictors (i.e., the set of all possible predictor combinations). Mapping this to our domain, the p predictors are the five components of the state plus action taken by Agent 1 at that state. Hence each row of the X matrix (corresponding to state s_1^t) would be of the form $\langle \tau_2, p_1^{t-1}, p_2^{t-1}, a_2^t, a_2^{t-1}, a_1^t \rangle$. The Y matrix corresponds to the future observation, i.e., the action Agent 2 takes at the next time step a_2^{t+1}. The X′ matrix has the same form as the X matrix and is obtained as the following Cartesian product: $n' = |\tau_2| * |p_1^{t-1}| * |p_2^{t-1}| * |a_2^t| * |a_2^{t-1}| * |a_1^t|$. We use a ridge estimator for purposes of regularization i.e., avoid overfitting by penalizing large coefficients (Witten & Frank, 2005).

4.1 Input Generation for Regression

We will now describe the training input generation for the logistic regression. Figure 4 shows a simple negotiation transcript that transpires between the agents. The figure shows three columns. Column 1 corresponds to the turn number, column 2 the player currently negotiating, column 3 corresponds to the action code while column 4 corresponds to the dialogue of the player in that round. We assume that we have labeled transcripts i.e., both the player types are known to us. Given this information, we now need to generate the input files for the logistic regression.

The first step in this process is to convert the negotiation transcript into a sequence of states. For example in Fig. 4, we start with an initial POMDP state with the following state vector: $< 0, 100, 0, a_1, NULL >$. Given that the transcript is labeled we know that $A2$'s type is 0 here. At the start of the game $A1$ wants a price of 100 and $A2$ wants a price of 0. The past code at the start is NULL since no code is seen till then. The future code is the action that $A2$ takes immediately following $A1$'s action, i.e., the observation for $A1$. The next state is constructed again for $A1$'s action. Therefore if at step 1, $A1$ takes an action and the transcript is strictly sequential the states are constructed at steps 1,3,5,... n (corresponding to times 0,1,2 etc., i.e., $t = ((step\%2) - 1)$ for a transcript with n steps. At step 3 the new tuple looks as follows: $< 0, 100, 0, a_7, a_1 >$. $A2$'s type remains same since the type is fixed. The proposal on the table remains the same since no proposal actions were taken in steps 1 and 2. $A2$ can take any of the 11 actions in the next time step. In this transcript, $A2$'s action at step 4 is a_7 and hence captured as the future action in the new tuple. $A2$'s action at step 2 was a_1 and hence the previous action was set to a_1 in the new tuple. At steps 47 and 50 the components 2 and 3 of the state corresponding to the previous proposal on the table change since the players take proposal actions.

Once the sequence of states are obtained, the setting of inputs (generation of training file) for the logistic regression is pretty easy. We match each of the odd numbered step s to t and form the rows of the X vector. $A2$'s action at step $s + 1$ becomes the corresponding row in the Y matrix. The X′ matrix is obtained as a Cartesian product of all the possible values for all the factors. The matrices are then passed to the WEKA logistic regression solver in the appropriate format.

Turn	Speaker	Code	Dialogue
1	A1	a_1	Hi, shall we start with price
2	A2	a_1	Sure
3	A1	a_2	Do you like mid-way price?
4	A2	a_7	Not really
.			
.			
.			
47	A1	a_4	My best price is 40
48	A2	a_2	Is there something else we can do?
49	A1	a_1	Hmm...
50	A2	a_7	Fine. Let's go with it.
.			
.			
.			

Fig. 4 A sample negotiation transcript

The WEKA solver first builds a classifier using the input X and Y matrices. The classifier then takes as input each instance from the X′ matrix and computes a distribution for that instance. This distribution corresponds to the predicted observation in the next state and hence determines the transition probability between the current state and agent action to the new state.

5 Experimental Results

5.1 Generation of Real Negotiation Data

As mentioned earlier, we use pre-existing inter-cultural real negotiation Cartoon data sets (Adair & Brett, 2005) for our experiments. As described in Adair & Brett, the scenario was played out between different participants on a one-on-one basis.

Fig. 5 Effect of increasing order on predicting the next code
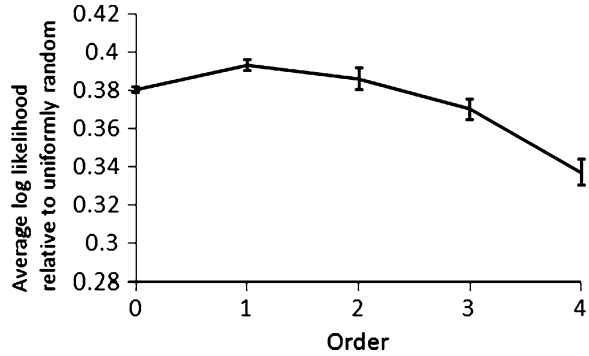
Each of the two participants were from a different culture and were given a 90-min deadline to reach a negotiation after they become comfortable with the scenario. For purposes of our experiments we use US versus four other cultural dyads. These four other cultures are: (a) Japanese (b) Israeli (c) German and (d) Chinese. We used a total of 27 US-Japanese, 26 US-Israeli, 20 US-German and 13 US-Chinese dyads in our experiments.

5.2 Predicting Next Code

Before we start analyzing the quality of the transition function generated, it is important to understand whether there will be any benefit of using the *first* order or in general a *n*th order POMDP as opposed to a *zero* order model. This is important to investigate since the size of the POMDP state space grows exponentially with n and the problem becomes intractable. To demonstrate the usefulness of *n*th-order model, we compare the performances of various n order models at predicting the next observation (i.e., action of the opponent) given the previous n observations. For example, for a n of 0, we try to predict the next observation without using any history while for a n of 1 we predict the next observation using one previous observation of the opponent. In general for a given n we predict the next observation using the n previous observations.

Figure 5 shows the order (n) on x-axis and the prediction capability captured as a average log likelihood relative to the uniform random policy on the y-axis. The experiment has been performed on real negotiation data involving the dyads mentioned earlier. From the graph we notice that a *first* order model (n = 1) has a better prediction capability over a *zero* order model. The result as n increases is surprising since we expect a better prediction capability for higher n's. However our results show that the prediction capability decreases with further increase in n. We attribute this to lack of sufficient data and further analysis would be needed to make general conclusions. However, for the current data set a *first* order model is the optimal choice for predicting the next code and hence used in the rest of our experiments.

Fig. 6 Pair-wise comparison test
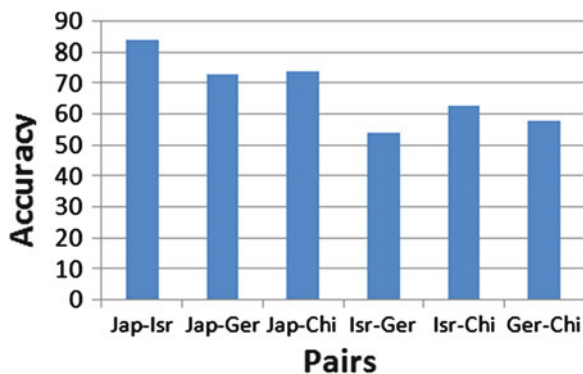
5.3 Predicting the Culture of the Opponent

The transition function of the POMDP captures the key behaviors of the opponent players and hence it is important for us to analyze whether it learnt the appropriate behaviors. We present here results to show that indeed the transition function learnt the correct behavior using the logistic regression. To demonstrate this we performed the belief tracking experiment i.e., we use only the belief update part of the POMDP, and ignore the reward and the optimized policy. This means that we pre-specify the actions of both A1 and A2 from the test dataset instead of allowing the POMDP to pick the optimal action. We then check if the POMDP is able to classify the agent type correctly using the POMDP belief updates. A high classification accuracy would imply that the POMDP was indeed able to learn the behaviors appropriately.

We set up the belief tracking experiment as follows: The POMDP is trained using the logistic regression as the US player while the opponents can be chosen in the following ways from the four possibilities i.e. Japanese, Israeli, Germans and Chinese.

– Pair-wise comparison between two cultures (ex: Japanese vs Israeli, Japanese vs Germans etc) leading to a total of six different experiment settings
– One culture against rest of the cultures combined together (ex: Japanese vs Israeli/Germans/Chinese combined together) leading to a total of four different experimental settings

Figure 6 shows the classification accuracy for the pair-wise testing experiment. The x-axis of the figure lists the various pairs used as opponents while the y-axis shows the classification accuracy i.e. number of times the belief-update procedure classifies an opponent from a particular culture as an opponent from that culture with greater than 0.6 (60%) belief in that classification. Note that initially the POMDP believes that the opponent can be from one of the cultures with a 0.5 (50%) belief. We classify the results as inconclusive, for the belief range between 0.5 and 0.6 and hence not counted towards correct classification. From the figure we observe that the

Table 1 Results of belief
tracking experiment

Culture	Accuracy
Japanese	84
Israeli	64
German	54
Chinese	59

Japanese culture can be distinguished with a high accuracy from the other cultures in a pair-wise comparison test with >70% accuracy in all the tests. The Chinese culture seems to be the next best in terms of classification accuracy with >60% accuracy in all the tests. While distinguishing between Germans and Israeli opponents has been hard in our tests, even this worst case classification accuracy has been greater than 50% (i.e., a random classification), thus showing that a logistic regression based classification is indeed useful in distinguishing between the opponents from different cultures in general real-world settings. The results get better once we start considering instances as correctly classified when the belief of the POMDP is >50% (as opposed to >60% till now). For example, there is a decisive swing of 9% increase in accuracy for the worst opponent pair i.e. Israeli vs Germans with this relaxation.

Our next experiment involves trying to distinguish between opponents where one opponent is from a single culture while the other opponent is the rest of the cultures treated as a single opponent. The idea here is to see if this additional cueing may help in obtaining a better accuracy. Table 1 shows the classification accuracy of this experiment. As can be seen from the table the classification accuracy is in general higher than a pair-wise classification experiment. For example, the Japanese culture can be distinguished with 84% accuracy when the rest of the cultures are treated as a single opponent as opposed to the worst case accuracy of a little more than 70% during a pair-wise classification experiment. Similar trends hold for the other cultures too with smaller improvements in classification accuracy. As seen earlier, the accuracy results get better if the threshold used for classifying as an accurate prediction was reduced to 50% from the 60% used here. We therefore conclude that the POMDP model was able to leverage the extra cueing that was present in the data due to the combination of various cultures and was able to perform much better than a random classification (50% accuracy).

Finally, Fig. 7 shows a snapshot of the average probability difference across the action set for a Japanese vs rest of the other cultures combined together as an opponent. In particular, the x-axis of the figure shows the 11 actions that the POMDP can take while the y-axis shows the averaged difference in probability across all the states for the two different cultures. The calculation for this is done as follows: There are a total of 27,000 states in the POMDP with the first 13,500 states (top-half) pertaining to Japanese culture (type = 0) while the rest of the states (bottom-half) represent the rest of the cultures combined together. For each possible action, we take the mean probability across all states for that action for the Japanese culture and subtract the mean probability across all states for the other opponent. The graph shows that except for actions 1 and 7 the average differences are negligible. Action
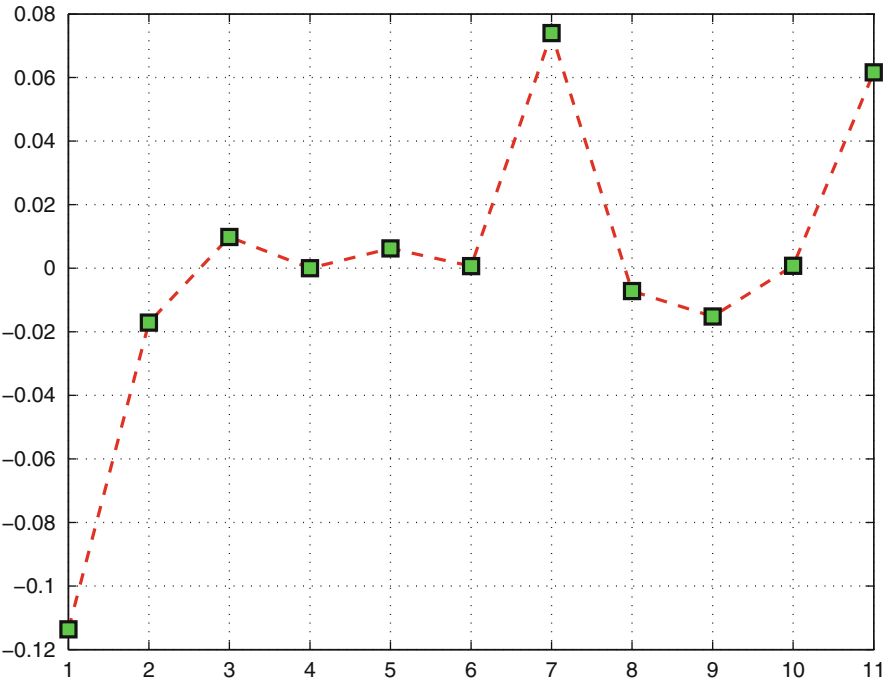
Fig. 7 Average action probability difference across various cultures

1 corresponds to information action and action 7 corresponds to reaction. Based on this we conclude that actions 1 and 7 play a key role in which the Japanese culture could be distinguished from its opponent with 84% accuracy (see Table 1).

6 Related Work

There is a vast body of research for identifying the relevant psychological factors and building a theory of negotiation (Pruitt, 1981; Weingart, Brett, Olekalns, & Smith, 2007). We will briefly discuss here the literature that considers cultural effects in negotiation (for a more in-depth discussion see Gelfand & Brett, 2004). The effect of the cultural background of the negotiators on the negotiation process and the negotiation outcome has been studied both theoretically and experimentally in the behavioral sciences literature (see Brett & Okamura, 1998; Gelfand & Brett, 2004; Tinsley, 1998). Cultural values and norms affect the importance people ascribe to different issues and their interpretation of the opponent's behavior. Brett (2000) identifies and discusses the effect of three cultural values in cross-cultural negotiation: (a) individualism versus collectivism (b) egalitarianism versus hierarchy, and (c) high versus low context communication. There is a cultural

stereotype between the East and the West based on these values. A typical Western individual is presumed to be individualistic, egalitarian and uses low context communication whereas a typical Eastern individual is collectivist, hierarchical, and uses high-context communication. However, this is a very gross characterization and cultural differences within neighboring regions also affect negotiation ((Leung & Tjosvold, 1998) discusses this in the context of East Asian cultures). Although the knowledge of the opponents culture may be helpful in negotiation, there has been studies showing that negotiations can break down when negotiators adjust to their opponent's culture and try to overcompensate (Adair, Taylor, & Tinsley, 2009). Another important cultural factor in negotiation is cultural *sacred values* of the negotiators. People have high emotional attachment to the sacred values (Atran & Axelrod, 2008) and any act during the negotiation process perceived to violate them may result in a breakdown of the process.

There has also been some effort into building computational models of negotiation and building software agents for negotiation (see Braun et al., 2007; Kraus, 2001 and references therein). However, there is relatively little work on including the effect of culture in the computational models of negotiation (except Hofstede, Jonker, & Verwaart, 2008). The computational models for negotiation use a variety of techniques from game theory (Napel, 2002; Camerer, 2003), probabilistic decision theory (von Neumann & Morgenstern, 1944), Bayesian learning (Zeng & Sycara, 1998), and other heuristic approaches (Faratin, Sierra, & Jennings, 1998). It has been noted that the way people act in a negotiation (or in general strategic interaction) scenario do not always correspond to a game-theoretic equilibria (Camerer, 2003; Kagel & Roth, 1995). Moreover, many game-theoretic and decision-theoretic models have been mainly interested in the outcome of the negotiation game instead of process of negotiation. We are crucially interested in both outcome and process in modeling cultural effects in negotiation. That said, newer game-theoretic analyses do pay attention to the process as well as the outcome; a full comparison of statistical, decision theoretic, and game theoretic techniques is an open research problem. POMDPs have been used before in modeling human social interaction where knowledge of the opponents needs to be maintained (Pynadath & Marsella, 2005). More recently, POMDPs have been used in a game-theoretic setting for modeling a finite repeated game between two agents (Ray, King-Casas, Montague, & Dayan, 2008).

7   Conclusions and Future Work

In this chapter, we first showed the mapping of a general cognitive schema for negotiation to a POMDP model. We then described the POMDP setup for a general negotiation problem and discussed the challenges in modeling a negotiation problem using POMDPs. We presented a simple example of a single-issue transactional negotiation to illustrate the POMDP formulation. While many competing techniques to model a negotiation problem exist, the POMDP based modeling has the following

advantages: (a) POMDPs provide a decentralized way of solving the problem which is an inherent characteristic of the negotiation domain. (b) POMDPs provide a natural way to capture the sequential nature of the bargaining process, i.e., they capture the process rather than just focusing on the outcome. (c) POMDPs can express the various important factors that affect the negotiation such as culture. We also discussed how cultural factors can be accounted for and how they affect the POMDP modeling. We then presented results using real data to support the fact that a POMDP based modeling is indeed effective.

References

Adair, W., & Brett, J. M. (2005). The negotiation dance: Time, culture, and behavioral sequences in negotiations. In *Organization Science, 16*(1), 33–51.

Adair, W. L., Taylor, M. S., & Tinsley, C. H. (2009) Starting out on the right foot: Negotiation schemas when cultures collide. *Negotiation and Conflict Management Research, 2*(2), 138–163.

Atran, S., & Axelrod, R. (2008, July) Reframing sacred values. *Negotiation Journal, 24*(3), 221–246.

Braun, P., Brzostowski, J., Kersten, G., Kim, B., Kowalczyk, R., Strecker, S., & Vahidov, R. (2007). e-negotiation systems and software agents: Methods, models, and applications. In *Intelligent decision-making support systems* (pp. 271–300). London: Springer.

Brett, J. M. (2000). Culture and negotiation. *International Journal of Psychology, 35*(2), 97–104.

Brett, J. M., & Okamura, T. (1998). Inter- and intra-cultural negotiations: US and Japanese negotiators. *Academy of Management Journal, 41*, 495–510.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interactions*. Princeton, NJ: Princeton University Press.

Drake, L. E. (2001). The culture-negotiation link: Integrative and distributive bargaining through an intercultural communication lens. *Human Communication Research, 27*(3), 317–349.

Druckman, D. (1986). Stages, turning points and crises: Negotiating military base rights, Spain and the United States. *The Journal of Conflict Resolution, 30*(2), 327–360.

Faratin, P., Sierra, C., & Jennings, N. R. (1998). Negotiation decision functions for autonomous agents. *International Journal of Robotics and Autonomous Systems, 24*(3–4), 159–182.

Gelfand, M. J., & Brett, J. M. (Eds.). (2004). *The handbook of negotiation and culture: Theoretical advances and cultural perspectives*. Stanford, CA: Stanford Business Books.

Hilbe, J. M. (2009). *Logistic regression models*. Boca Raton, FL: CRC Press.

Hofstede, G. J., Jonker, C. M., & Verwaart, D. (2008). An agent model for the influence of culture on bargaining. In *HUCOM 2008, first international working conference on human factors and computational models in negotiation* (pp. 39–46). Delft: Netherlands.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*, 99–134.

Kagel, J., & Roth, A. (Eds.). (1995). *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Kraus, S. (1997, July). Negotiation and cooperation in multi-agent environments. *Artificial Intelligence, 92*(1–2), 79–97.

Kraus, S. (2001). *Strategic negotiation in multiagent environments*. Cambridge, MA: MIT Press.

Leung, K., & Tjosvold, D. (Eds.). (1998). *Conflict management in the Asia Pacific*. Singapore: Wiley.

Myerson, R. B. (1977). Two-person bargaining problems and comparable utility. *Econometrica, 45*(7), 1631–1637.

Napel, S. (2002). *Bilateral bargaining: theory and applications*. Berlin, The Netherlands: Springer.

Pineau, J., Gordon, G., & Thrun, S. (2003). Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI*. Acapulco: Mexico.

Pruitt, D. G. (1981). *Negotiation behavior*. New York: Academic Press.

Pynadath, D. V., & Marsella, S. C. (2005). PsychSim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the international joint conference on artificial intelligence* (pp. 1181–1186). Edinburgh: Scotland.

Ray, D., King-Casas, B., Montague, P. R., & Dayan, P. (2008). Bayesian model of behaviour in economic games. In *NIPS*. Vancouver: Canada.

Rubinstein, A. (1982, January). Perfect equilibrium in a bargaining model. *Econometrica, 50*(1), 97–109.

Tenbrunsel, A. E., & Bazerman, M. H. (1995). Working women. In *Teaching manual*. Evanston, IL: Dispute Resolution Research Center, Northwestern University.

Tinsley, C. H. (1998). Models of conflict resolution in Japanese, German, and American cultures. *Journal of Applied Psychology, 83*, 316–323.

Tinsley, C. H. (2001). How negotiators get to yes: Predicting the constellation of conflict management strategies used across cultures. *Journal of Applied Psychology, 86*(4), 583–593.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior* (2nd ed., 1947 edition). Princeton, NJ: Princeton University Press.

Weingart, L. R., Brett, J. M., Olekalns, M., & Smith, P. L. (2007). Conflicting social motives in negotiating groups. *Journal of Personality and Social Psychology, 93*(6), 994–1010.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Boston, MA: Morgan Kaufman.

Zeng, D., & Sycara, K. (1998, January). Bayesian learning in negotiation. *International Journal of Human-Computer Studies, 48*(1), 125–141.

Modeling Dynamics of Multicultural Integration and Conflict

Wouter E. de Raad, Andrzej Nowak, and Wojciech Borkowski

Abstract A series of computer simulations explored a formalized theoretical model of multicultural integration which concentrates on attitudes towards ingroup and outgroup as main variables. The simulations were aimed at the investigation of social dynamics underlying the evolution of patterns of integration and segregation. The findings suggest that: integration is possible only with strong mutually positive attitudes; negative attitudes lead to segregation even if one group would be positive; surprisingly tolerance – a neutral stance also leads to segregation. These results are discussed in the light of conflict and multiculturalism.

Keywords Conflict • Culture • Multiculturalism • Multicultural integration • Segregation • Computer simulation • Dynamical model

W.E. de Raad (✉)
Faculty of Psychology, University of Warsaw, Warsaw, Poland

Institute for Social Studies, University of Warsaw, Warsaw, Poland
e-mail: wderaad@psych.uw.edu.pl

A. Nowak
Warsaw School of Humanities and Social Sciences, Florida Atlantic University,
Boca Raton, FL, USA

Institute for Social Studies, University of Warsaw, Warsaw, Poland

W. Borkowski
Institute for Social Studies, University of Warsaw, Warsaw, Poland

Warsaw School of Humanities and Social Sciences, University of Warsaw, Warsaw, Poland

1 Introduction

One of the major challenges of our time concerns the peaceful coexistence of peoples from different cultures. The world stage provides us with ample evidence that cultural differences can and do lie at the basis of conflict. One important aspect of this issue is related to migration and immigration and the increasingly multicultural make-up of societies (Castles & Miller, 2009). The fact that societies often struggle with issues related to multiculturalism seems to be the rule rather than the exception. The popularity of political parties that base their views on multicultural intolerance designates how widespread and pervasive people's concern with multiculturalism is (Rydgren, 2007). But also in situations where migration issues do not play a large role we can observe conflicts on the basis of cultural differences. Examples of outbursts of violence in Iraq, India, Pakistan, Sri Lanka, Nepal, Philippines, Sudan, Somalia, and many Arabic countries (to name some examples that have attracted considerable international attention) all seem to depend on differences of values and goals of the parties involved. These conflicts may be rooted in differences of religion, political ideology, ethnicity, or history, but at a more general level it seems that those conflicts are related, at least in part, to differences in culture. The "objective" cultural differences are obviously of less importance than the perceived differences by the parties in conflict. In this respect it may be fruitful to use a rather broad understanding of what constitutes cultural differences and also look at differences between subcultures within an ethno-culturally homogeneous group. The fact that the large majority of the current violent conflicts in the world takes place within the borders of a nation state (Harbom & Wallensteen, 2009) supports this idea.

Since most current conflicts are intra-national, peace efforts, instead of being aimed at bringing peace between countries, have to focus on bringing peace within countries. This has implications for the kind of initiatives that may result to be effective. Peace cannot be brought about by establishing, respecting or controlling borders, because citizens can often travel freely throughout their country. Moreover we are not dealing with two populations separated by borders at all, but with groups that often share the same land, space and institutions. Peace then de facto means "peaceful coexistence". As a consequence there is a very large social component to the understanding and resolution of conflict in these situations. Research on the relationship between social capital and conflict shows the importance of social contacts, patterns of interaction, and changes in those patterns (Colletta & Cullen, 2000; Putnam, 1993). The key issue then is to gain insight in the dynamic aspects of the social interactions that underlie the intensification or pacification of conflict.

The question is how we can theoretically capture this. The social processes implied involve the evolution of multi-agent interactions over time. Social psychological theories generally specify fixed-in-time causal relationships and seem unlikely to offer much insight in dynamic social processes. In this paper we will work towards a theory that allows us to understand the dynamic social aspects of conflict. A series of computer simulations explored a formalized theoretical model

of multicultural integration which concentrates on attitudes towards ingroup and outgroup as main variables. The simulations were aimed at the investigation of social dynamics underlying the evolution of patterns of integration and segregation. The findings suggest that integration is possible only with strong mutually positive attitudes; negative attitudes lead to segregation even if one group would be positive. Surprisingly, tolerance – a neutral stance also leads to segregation. These results are discussed in the light of conflict and multiculturalism.

2   Computer Simulations of Multicultural Integration and Conflict

It has been convincingly argued that computer simulation is a valuable tool for theory construction in social psychology (Gilbert & Troitzsch, 2005; Ostrom, 1988; Vallacher & Nowak, 1997). They are especially useful in uncovering the dynamics and finding which parameters play a crucial role in these systems. Computer simulations allow one to generate predictions and hypotheses on the basis of formalized theoretical notions and thus offer a big advantage in the formulation of theory of complex dynamic social systems. In a simulation a modeled complex and lengthy real-time social process can be visualized and studied in an instant of time. Systematic variation of variables allows investigating the relationships between parameters, dynamics and outcomes of social interactions that would be hard to observe in reality. Additionally, and important in the light of conflict, simulations may help us to understand the effect of interventions in a complex social system. One obstacle to theory development in this way is that current social theories often lack the precision to formulate precise rules that can be used in a computer simulation. However, existing theories and models can often be adapted, or can serve as inspiration for simulation studies.

Schelling (1971) has shown that a simple rule: "If an individual is surrounded by more individuals of different type than the number of individuals of own type, then the individual moves from the current location to a random empty location", leads to complete segregation. Although Schelling developed his model to explain racial segregation the results can be adopted to explain segregation of people from different cultures. Although the model of Schelling explains why individuals from different cultures may live in separation, it does not explain why it is rarely the case, and how contacts between different cultures may depend on psychological and sociological factors.

Berry's model of acculturation (1980, 1997), although not directly related to conflict, describes how groups of people of different ethno-cultural background adapt and choose to live together in culturally plural societies. According to Berry people have different views about how they want to live together and accordingly they adopt different "acculturation strategies". When individuals do not wish to maintain their cultural identity, but prefer to maintain contacts with the other culture, this is called the assimilation strategy. On the other hand, if individuals prefer to

hold on to their original culture and want to avoid contact with the other group, this is called the separation strategy. When an individual is both interested in maintaining one's original culture as well as seeking interaction with the other culture, integration is defined. When in contrast there is neither interest in cultural maintenance nor in cross-cultural interaction, marginalization is defined. Based on this model Bourhis and colleagues (Bourhis, Moïse, Perreault & Senécal, 1997) proposed an interactive model of acculturation and predicted different relational outcomes (consensual, problematic and conflictual) based on the combination of the acculturation strategies of the groups in contact. Their predictions are based on expected consequences of combinations of acculturation strategies, not on observed interactions between groups or individuals. So although the interactive model of acculturation is not dynamic in nature, it shows that Berry's model lends itself to make inferences about interactions and outcomes.

In an entirely different approach Mimkes uses binary alloys as a model for the multicultural society (Mimkes, 1995). His model is borrowed from thermodynamics and is truly dynamic in nature; the final state of the system depends on the interactions of the agents (or particles). The behaviors of people in this model are specified by applying the same limited set of rules that apply to physical particles and he outlines striking similarities between the results on molecular level (solubility of molecules) and societal level (patterns of intermarriage between races or religions). Despite the parallels, the precision of the model and its elegant simplicity, a fundamental problem is that we know that people behave in a different fashion than molecules. That is, even though the outcomes of this model may be comparable to societal situations, the dynamics that produce these outcomes do not compare to social dynamics and are unlikely to provide the understanding that we seek. What we need is a model that combines the precision and dynamic nature of Mimkes' model, but is firmly based on empirically supported social psychological theory, like Berry's model.

In this paper we develop a theoretical model that is based on Berry's model of acculturation but is larger in scope so that it captures a wider range of situations. We will formalize this model in precise rules so that it can be implemented in a computer simulation. We will analyze and discuss the outcomes of a series of computer simulation in the light of conflict and multicultural integration. More specifically we will look at several distinct scenarios that can be identified in the simulations and discuss their relevance to real life situations.

3 The Model

Berry's model as specified by its defining terms only applies to settings in which different ethno-cultural groups are in continuous contact (Berry, 1980, 1997). We want to broaden the scope of our model so as to include situations in which different groups interact that would be regarded as having the same ethnic background, but that differ according to lines of religion, politics or values. In Berry's terms those

groups may be regarded as different ethno-subcultures. Our model then may rather deal with multicultural integration than acculturation which would imply ethno-cultural differences.

In his model Berry makes a distinction between dominant/majority groups and minority groups. Acculturation strategies of minority groups are defined by their attitudes towards culture maintenance and contact with the majority group. The acculturation strategies of the majority group are defined in relation to the acculturation of the minority group however. For instance, if a person from the majority group prefers minorities to integrate, this acculturation strategy is termed multiculturalism. Thus, the acculturation strategies of the majority do not include an evaluation of their own culture, but are based on a preference for contact with minorities and a preference for minorities to retain their culture or not. This conceptualization well fits situations of relative recent migration in which a majority/dominant/mainstream culture can be identified and one or more minority groups. In some or even many cases this distinction might not be so clear however. Also the fact that only the minority group evaluates the question to retain its own culture is limiting and makes the model unbalanced.

A strong aspect of Berry's model is its simplicity: Attitudes are the main variables that inform peoples' decisions and behaviors that shape the relations between groups. For our model we assume that each person in a multicultural setting – whether minority or majority has an attitude regarding the importance placed on engaging in contact with people from the other culture and with people from the own culture. We regard these attitudes as continuous dimension ranging from positive to negative. Because humans are social animals and have an innate drive for social contact, we further posit that those attitudes determine with whom contact is sought and avoided so as to maximize contentment or happiness. These are the basic building blocks needed for a computer simulation.

4 Formalizing the Model

In order to simulate the model the relationships between the variables need to be specified. We posited that people seek social contact with others to maximize their happiness; happiness therefore is a function of the number of social contacts. How contacts with others are valued depends on a person's attitude towards the group to which that person belongs: A positive attitude means that contact with that person increases happiness, a negative attitude is related to a decrease. Although happiness increases with the number of positively valued contacts, this increase is not linear. We used the Dynamic Theory of Social Impact (Lewenstein, Nowak, & Latane, 1993; Nowak, Szamrej, & Latane, 1990) to specify how happiness depends on the contact with others. The theory states that the impact of a group of individuals changes as a square root of the number of individuals exerting impact. This means that the increase in happiness for each additional contact becomes smaller. In other words, the effect on happiness of the first contact is larger than of the second,

which in turn is larger than the effect of the third, etc.[1] Because the focus lies on interactions with members from different groups, contacts from each group are counted separately. The relationship between those variables can be summarized with the following formula:

$$H = A_{own} * \sqrt{N_{own}} + A_{other} * \sqrt{N_{other}}$$

Explained in words, this formula expresses that happiness H depends on the sum of the product of one's attitude towards members of the own group A_{own} with the square root of the number of own group contacts N_{own} and the same product for members of the other group.[2] The decreasing marginal utility is implemented here by taking the square root of the number of contacts for each group.

Social interactions were simulated on a square lattice (a very large checker board): A two dimensional space that consists of cells. Individuals, or agents, were assigned to a group, and given an attitude towards the own and other group ranging from -1 to $+1$. Next they were located on a random spot or cell on the board. Some cells would remain vacant to allow for freedom of movement for the agents. Agents would be randomly selected and asked to evaluate their current happiness according to the formula specified above. They would then be offered random vacant locations to move to if desired. A decision to move would be made if a spot would provide a higher level of happiness than the current location. The simulation would end if no agent would move any longer or after an arbitrary set limit if agents would continue to move.

In their evaluations agents would take all directly neighboring cells into account; 8 in total. Empty cells would have no impact on the evaluation. A "cost" of moving was included by setting a rule that agents would only move if they could obtain at least a 10% increase in happiness. To ensure that all locations had an equal amount of neighboring cells, the board was turned into a torus, which visually resembles a donut-shape.

This type of simulation belongs to a class of models called cellular automata, which are particularly well suited for simulating dynamic social processes with locally interacting agents (Gilbert & Troitzsch, 2005; Hegselmann, 1998; Schelling, 1971) A large advantage of cellular automata is the ease with which they can be visualized and in that way observed and analyzed.

5 Simulations

The simulations are represented as a square grid consisting of 50 rows and columns. Twenty-five percent of the 2,500 cells were left empty, to provide ample of space for agents to move. The two interacting groups were created of equal size. At the

---

[1]Provided that attitudes towards those contacts are identical.

[2]For the sake of simplicity we limit ourselves to two interacting groups.

Fig. 1 Initial state of a
simulation in which agents of
two groups – depicted in *two
shades of gray* are randomly
distributed on a grid. *Black
color* indicates empty spaces
to which agents can
potentially move
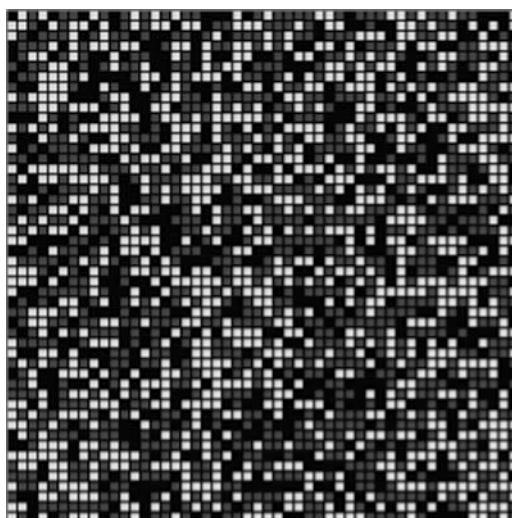
Fig. 2 Pattern of intergroup
contact between agents with
mutual negative attitudes
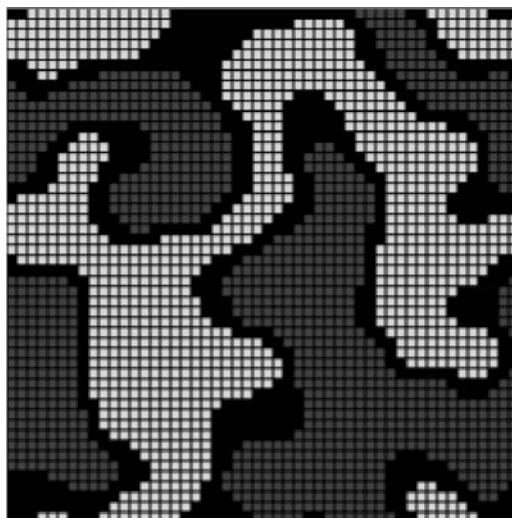
onset of a simulation agents were randomly assigned a location on the grid, which
would typically look like Fig. 1.

In the following simulations agents' attitudes towards their own group was
always assumed to be positive and was set at value 1.00. Although different values
could be easily implemented a positive ingroup attitude seems to make the most
sense in the light of most situations. Figure 2 depicts the outcome of a simulation in
which groups had mutually negative attitudes. The initial state of this simulation
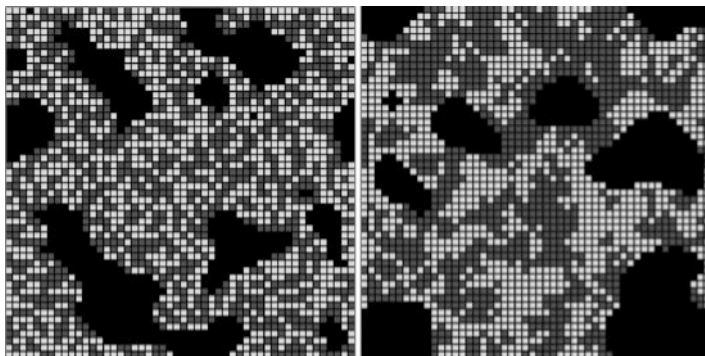
Fig. 3 Pattern of intergroup contact between agents with equal and mutually positive attitudes towards each other. In the picture on the *left* mutual attitudes are 1.00; in the *right picture* the mutual attitudes are 0.50

looked like Fig. 1. Next, agents were allowed to move according to the rules specified above until, in this case, the simulation stabilized and agents did no longer move.

The groups are completely segregated from each other, with not a single instant of contact between them. Simulations with different negative attitude values show that the resulting patterns are the same regardless of how negative the attitudes are: Negative mutual attitudes lead to complete segregation. Although negative attitudes per se do not indicate conflict or violence, it is clear that a situation that is characterized by negative mutual attitudes and lack of contact between groups is at large risk of developing into conflict and potentially violence. Varshney (2001) in detail shows how the absence of contact and communication is related to violent conflict. He compares how Hindus and Muslims in different cities in India reacted to equally violent-prone situations. Because of a relative absence of local interactions between the Hindus and Muslims in some places – both on a day to day basis and on an associative level there were no mechanisms that could successfully contain or stop violence.

In cases where people have strong positive mutual attitudes the situation might look similar to left side of Fig. 3. In this simulation agents had equally strong attitudes towards the other group as towards the own group. The resulting pattern is entirely random; group membership does not have an influence on the pattern of contact. This may be compared to people interacting in the United States who are of Italian or Irish descent; it may simply not matter (anymore) to which group one belongs. Another example to think of is the interaction between Catholics and Protestants in the Netherlands. The fact that most people don't know that the southern part of the Netherlands is predominantly catholic and the north mainly protestant illustrates that both groups integrate so well that this particular make up of the society hardly ever receives any attention.

The observations that strong mutual negative attitudes lead to segregation and strong mutual positive attitudes lead to integration may not be very surprising. The
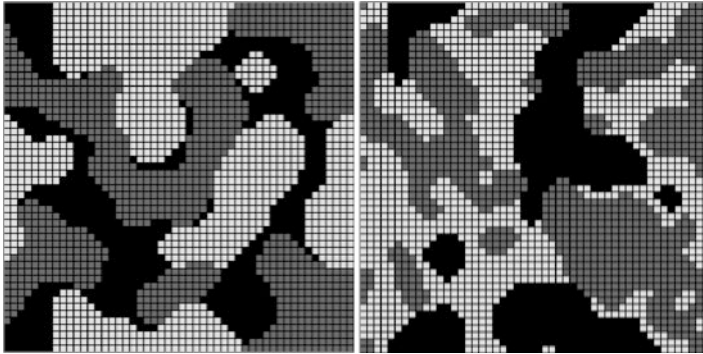
Fig. 4 On the *left*: A pattern of intergroup contact between mutually tolerant agents; attitude towards the other group is 0.00. On the *right* the *lighter shaded* group has a moderately positive attitude of 0.50, the *dark shaded* group has a neutral attitude of 0.00

simulations show however that also mild mutual negative lead to total segregation. A question then is if a pattern of integration can also be observed for any level of positive mutual attitudes. The right side of Fig. 3 depicts the result of a simulation in which the groups had moderately positive attitude: 0.50. Still both groups are integrated with each other, but it is clear that some extent of clustering has taken place. Although the groups are not segregated, the pattern is not a random one like in the left hand picture. The explanation for this is that agents are more attracted to people from their own group than to people from the other group. Still this situation represents a high degree of contact between the groups and may be comparable to a situation in which people place importance on their group membership and maintain their culture by participating in organizations, events, customs, rituals, etc., but at the same time maintain contacts with the other groups in social life, at work, in formal and social organizations, etc. It is especially this intercultural form of social capital which is based on social ties between the groups that is so important in the prevention and resolution of conflict (Varshney, 2001).

In public debates in many countries with a culturally plural population tolerance is often touted as the key to successful multicultural integration. Tolerance, the acceptance of others who are different although one may not agree with them, can be represented in our model as having a neutral stance toward the other, or an attitude with value zero. Given the face validity of the claim that tolerance is instrumental in establishing integration, it was much to our surprise that we observed that our tolerant agents did not integrate as well as we expected. As can be seen in Fig. 4, tolerance seems to lead to segregation rather than integration.

Segregation in this case is not total, as it would be if mutual attitudes would be negative, still the result is striking. It is mainly at the group boundaries that zones of contact between the groups exist. It is of course positive that contact between the groups exists at all, but from these results it would be hard to conclude that tolerance automatically leads to integration. Even in the panel on the right, where one of the groups has a moderately positive attitude towards a tolerant group, there
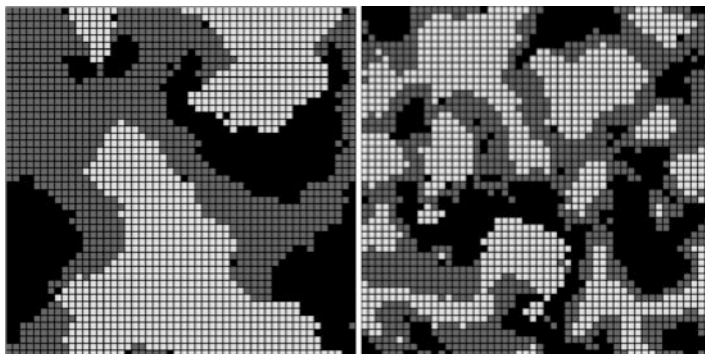
Fig. 5 In both pictures the *light colored* group has a mild negative attitude towards the *dark colored* group: −0.25. The *dark colored* group has a positive attitude towards the light group: 0.50 on the *left hand picture*, 0.75 on the *right*. These simulations do not result in a stable configuration, but display ongoing movement

are relatively few integrated pockets and most agents have no contact with the other group. The reason why this pattern emerges is that if people are neutral towards the other group, they will only seek contact with their own. They will not avoid contact with the other, but will not look for it either. In this case segregation is not the result of bad intent, but of indifference. This situation seems to characterize many real life situations; in many places ethnic communities exist that function autonomously from the larger society or from other groups to a certain extent. We can think of Poles in Chicago, Cubans in Miami, Turks in Berlin, or Algerians in Paris, for example.

A situation like the one sketched in the right panel of Fig. 4, in which two groups inhabit largely different worlds may balance on a tipping point of failure and success of integration. If the contacts that do exist help to establish more contact between the groups, then it is likely that this initial state of tolerance will develop into a more integrated whole. If however forces apply that drive the groups apart, and the relatively few contacts that exist diminish, it is not hard to imagine that a situation may occur that is susceptible to conflict or violence. So although tolerance is obviously to be preferred over disliking, it should be seen as the beginning of the road to successful integration and sustainable peace; as a prerequisite, rather than the solution.

Many contemporary Western multicultural societies are characterized by immigrant groups willing to integrate into the larger society, and thus having at least a moderately positive attitude towards the mainstream group (Berry, 1997). On the other hand the current political climate, informed by the popularity of parties with an anti immigration agenda, suggests that the larger population may often have a slightly negative attitude towards immigrant groups. Figure 5 shows two situations in which the light shaded group has a mildly negative attitude towards the darker shaded group (−0.25). On the right hand side the dark shaded group has a moderately positive attitude (0.50) toward the light shaded group; on the left hand side this attitude is 0.75.

What happens in these situations is that members of the dark shaded group will seek contact with the lighter shaded group. Because of their slightly negative attitude toward the other group, light shaded group members try to avoid contact with the other and will move away if approached. What results is what may be described as a situation of hide-and-seek, or a chase, in which one group constantly approaches, and the other group constantly withdraws. In an immigration setting this may resemble a situation in which immigrants are motivated to participate in the society through means of work, study or day to day contacts, but are met with general rejection. They will be accepted within their own community, but in the society at large they will not be able to successfully participate as a group, because as long as the other group actively avoids contact segregation will result.

In the situation on the left in Fig. 5 the dark shaded group has a stronger positive attitude towards the lighter group; 0.75, compared to 0.50 on the right. Because of their stronger positive attitude the dark shaded group more vigorously chases the light shaded group around. Judging the picture it may seem like this leads to a more integrated state, but remember that this is a snapshot of an ongoing process of movement.

These results are relevant in the light of many public debates relating to multicultural integration and the question that is sometimes coined: Who is responsible for integration? For members of the dominant group – the mainstream society it is easy to say that the responsibility for multicultural integration lies at the feet of the immigrants. What these simulations teach us is that even if immigrants are really willing to integrate, even a relatively small negative attitude in the society at large may be preventing it from happening. Ironically, the more negative people are towards immigrants, the more emphasis they seem to place on the immigrants' responsibility to adapt or integrate. These results are very clear however in showing that responsibility for preventing integration lies at the feet of only one group. The fact that, especially in the picture on the left, there is still a considerable amount of contact between the groups is because the agents have a relentless motivation and drive to get in contact with the other, irrespective of the rejections. This clearly is unlikely be to be the case in reality.

In case the attitude of one of the groups (light) towards the other (dark) is very negative (−1.00), as in Fig. 6, we can see that even if the other group has a slight positive attitude of 0.25, a state of complete segregation results. The light shaded group has such a strong aversion to the other group that they ensure to avoid any contact. The attraction of the dark shaded group is not strong enough to make them seek contact with the members of the other group at the expense of having to give up contact with some of their ingroup members: Inspection of the picture shows that for a dark shaded agent to move to the edge, where contact with the other group is possible, would automatically mean to reduce the number of ingroup contacts. In this case a parallel can be drawn with the way in with people often have to negotiate the extent to which they identify with one group or another. In a case like this it seems that the question is either/or, and any sort of dual group membership is obstructed by the negative stance of one group.
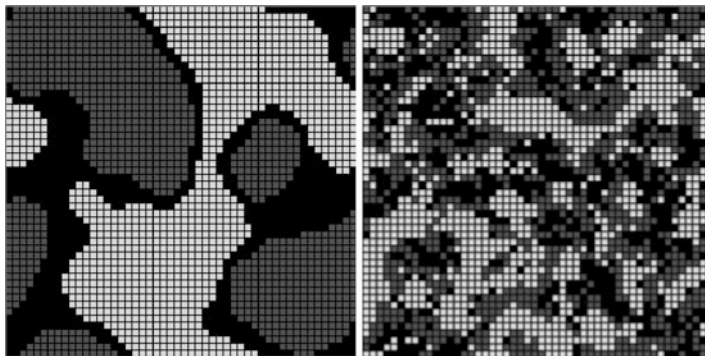
Fig. 6 In both pictures the *light colored* group has a strong negative attitude towards the *dark colored* group: −1.00. The *dark colored* group has a positive attitude towards the light group: 0.25 on the *left hand picture*, 0.50 on the *right*. The configuration on the *right side* is stable; agents no longer move. The configuration on the *left side* however is not stable and movement continues

In the picture on the right side the situation is comparable to the right side picture in Fig. 5; the attraction of the dark shaded individuals is now strong enough so that they will seek for contact even if this limits the number of ingroup contacts. Because of the strong wish to avoid contact by the light shaded agents again a chasing-game results that has no end to it.

6 Concluding Discussion

The results of our simulations suggest the existence of a limited number of scenarios of multicultural integration and provide insight in the development of these scenarios over time. In case the situation is characterized by mutual negative attitudes, any existing cross-cutting social tissue will likely disappear, leaving two segregated antagonistic groups. Without contact there will be no communication, which will make the development of mutual trust virtually impossible. Internal dynamics of the groups will more likely lead to escalation than to mitigation. Without cross cutting ties there are no longer mechanisms of containment that could help to pacify the situation; conflict is very likely. It is problematic that for any level of mutual negative attitudes segregation seems to follow. It of course makes a large difference if groups are only slightly negative or very negative towards each other, but once a movement of segregation has set in, internal group dynamics may propel the situation towards escalation. However, even out of the most hopeless situation peace can emerge as in Mozambique (Bartoli, Bui-Wrzosinska, & Nowak, 2010), and positive relations can deteriorate over time and end in violence, as in Rwanda (Colletta & Cullen, 2000). This should make us realize that in any situation the current state of the system should be interpreted in the light of the direction in which it is moving.

In situations where mutual sentiments are positive, groups may integrate. Unlike with mutual negative attitudes, where any degree leads to segregation, the extent of integration is closely related to degree of mutual positivity. With sustainable peace in mind, it is important to encourage improving attitudes in order to facilitate social integration. Intergroup contact theory suggests that there are mechanisms that are autocatalytic in improving attitudes through contact (Pettigrew, 1998), and even indirectly through other group members (Wrigth, Aron, McLaughlin-Volpe, & Ropp, 1997). The conditions under which these effects are found are limited unfortunately, and reality abounds in examples where contacts are more antagonizing than bringing people closer together. If mutual positive attitudes exist however there is a good climate for initiatives to strengthen contacts to be effective. Varshney's observations indicate that especially formal, organized forms of contact play an important role in the prevention of conflict. Small scale social contacts may be effective in ensuring local integration between groups, but institutionalized contacts seem to be much more effective in establishing the positive kind of relationships that are instrumental in controlling conflict on a larger scale (Varshney, 2001). Here it seems that a special role may be reserved for professional and religious organization as well as for the government. A comparison of several communities in Northern Ireland in the 1970s show that also leisure organizations where people from two sides meet may play an important role in preventing violent conflict (Darby, 1986).

Despite the popularity of the concept of tolerance, simulations with tolerant agents resulted in patterns of surprisingly strong segregation, even if one of the groups was trying to integrate into a tolerant one. When people are tolerant and indifferent towards contact with the other group and only actively seek ingroup contacts then segregation will be an unintended side effect. Preaching and fostering tolerance may well make sense in situations of antagonism, where reduction of conflict is the goal; it will by itself likely not lead to any significant extent of integration. In many societies where frictions between groups exist but no real antagonism, the best way forward is the promotion of social ties between groups.

Related to the observations above, even the best intentions from one group are not sufficient to force integration with the other. In a combination of groups with positive attitudes and with slightly negative attitudes some extent of contact is to be expected. If the attitudes of one group are very negative however, complete segregation may result even though the other group is mildly positive. Some simulations, in which one group has quite strong positive attitudes, show that despite negative attitudes of one side a considerable degree of integration is possible. However, these patterns are not stable, and this form of integration is not sustainable. It is only because the agents are not affected by each time they are rejected that this behavior does not stabilize, which is highly unlikely to happen in reality.

These simulations suggest that sustainable integration is only attainable with strong positive mutual attitudes between groups. Despite differing opinions regarding who is accountable for the lion's share of responsibility for integration, these simulations point out that a key role is reserved for any group or party that holds a less than positive attitude towards the other. In this light the promotion of tolerance

then may be seen as a remedy for conflict in the short term, it is the promotion of intercultural contacts that ensures peaceful coexistence in the long run.

Schelling (1971) has shown that the avoidance of neighbors of different type may lead to segregation. This study shows that on the basis of a simple theoretical model with a small set of assumptions computer simulations allow us to explore and identify various scenarios that provide insight in real life situations of multicultural conflict and peace. The underlying model is clearly an oversimplification of the reality as it presents groups (even literally) in a black and white manner. However this should be interpreted as a promise rather than a limitation. If the model would be extended to allow for individual differences between the agents, and for agents to be affected by the type of interactions they experience in the form of changing attitudes, for example, it would likely provide us with an even richer and deeper understanding of multicultural integration and conflict. This study shows the usefulness of computer simulations as the method of choice for anyone wanting to develop dynamic models and as a promising new method of inquiry in the field of conflict and culture.

References

Bartoli, A., Bui-Wrzosinska, L., & Nowak, A. (2010). Peace is in movement: A dynamical systems perspective on the emergence of peace in Mozambique. *Peace and Conflict: Journal of Peace Psychology, 16*, 211–230.

Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. Padilla (Ed.), *Acculturation: Theory, models and some new findings* (pp. 9–25). Boulder, CO: Westview.

Berry, J. W. (1997). Immigration, acculturation and, adaptation. *Applied Psychology: An International Review, 46*, 5–68.

Bourhis, R. Y., Moïse, L. C., Perreault, S., & Senécal, S. (1997). Towards an interactive acculturation model: A social psychological approach. *International Journal of Psychology, 32*, 369–386.

Castles, S., & Miller, M. (2009). *The age of migration: International population movements in the modern world* (4th ed.). Hampshire, UK: Palgrave Macmillan.

Colletta, N. J., & Cullen, M. L. (2000). *Violent conflict and the transformation of social capital*. Washington, DC: The World Bank.

Darby, J. (1986). *Intimidation and the control of conflict in Northern Ireland*. Syracuse, NY: Syracuse University Press.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist* (2nd ed.). Maidenhead, UK: Open University Press.

Harbom, L., & Wallensteen, P. (2009). Armed conflicts, 1946–2008. *Journal of Peace Research, 46*, 577–587.

Hegselmann, R. (1998). Modeling social dynamics by cellular automata. In W. G. Liebrand, A. Nowak, & R. Hegselmann (Eds.), *Computer modeling of social processes* (pp. 37–64). London: Sage Publications Ltd.

Lewenstein, M., Nowak, A., & Latane, B. (1993). Statistical mechanics of social impact. *Physical Review A, 45*, 763–776.

Mimkes, J. (1995). Binary alloys as a model for the multicultural society. *Journal of Thermal Analysis, 43*, 524–537.

Nowak, A., Szamrej, J., & Latane, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362–376.

Ostrom, T. M. (1988). Computer simulation: The third symbol system. *Journal of Experimental Social Psychology, 24*, 381–392.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.

Putnam, R. D. (1993). The prosperous community: Social capital and social life. *The American Prospect, 13*, 35–42.

Rydgren, J. (2007). The sociology of the radical right. *Annual Review of Sociology, 33*, 241–262.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Vallacher, R. R., & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry, 8*, 73–99.

Varshney, A. (2001). Ethnic conflict and civil society. *World Politics, 53*, 362–398.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73–90.

Index