Chapter 4

The Empirical Validation Of A Dynamic Systems Model Of Interaction; Do Children Of Different Sociometric Statuses Differ In Their Dyadic Play Interactions?

Abstract

Studying short-term dynamic processes and change mechanisms in interaction yields important knowledge that contributes to understanding long-term social development of children. In order to get a grip on this short-term dynamics of interaction processes, the authors made a simulation model of dyadic interaction of children during one play session, which is inspired by dynamic systems principles. The theoretical components of the model comprise children’s goal-directedness of actions, concerns, emotional appraisals, social power, and social effectiveness. The model’s output consists of simulations of children’s emotional expressions and actions over every second of a play session, of three groups of dyads of different sociometric statuses. This chapter describes the empirical validation of the model and the methods needed for such validation. It focuses on the model’s predictions of averages and distributions of the major variables and on the model’s sensitivity. Overall, the model fits the empirical data well. An exception is the lesser fit of the ‘popular’ group of dyads, which is explained by the limited use of social effectiveness in the model. In the discussion, we reflect - among others - on the implication of our findings for using this type of simulation models in the field of research on social development.

This chapter is based on: Steenbeek, H. & van Geert, P. (submitted to Developmental Science; currently under revision). The empirical validation of a dynamic systems model of interaction; do children of different sociometric statuses differ in their dyadic play interactions?
INTRODUCTION

Children develop a major part of their social behavioral repertoire in their interactions with peers (Rubin, Bukowski, & Parker, 1998). Important components in this developmental process are – among others - the child’s social skills, social competence, friendship relations, and sociometric status. Research often focuses on the possible predictive relationship between what children learn in their actual interactions with peers, and their long-term social functioning (Parker & Asher, 1987; Kupersmidt, Coie, & Dodge, 1990; Bagwell, Schmidt, Newcomb, & Bukowski, 2001; Prinstein & La Greca, 2004). An example is the relationship between having a rejected sociometric status and later dropout from school (Dixon, 1999; French & Conrad, 2001; Hymel, 1996; Kupersmidt & Coie, 1991). Our starting point is that the dynamics of long-term development are intrinsically related to the dynamics of short-term processes (Thelen & Smith, 1994; Lewis, 2002b). Therefore it is useful to study this dynamics of short-term social processes, in which child and context factors, such as the child’s social competence and sociometric status, play a central role. We constructed a general explanatory model of short-term social behavior in a developmental context. In this model, we use principles derived from dynamic systems theory on the one hand, and functionalist theory of emotions and theory of action were used on the other.

The model takes the form of a simulation model that generates interaction patterns of dyadic interaction of two children of different sociometric statuses during the course of one single play session. We especially focus on explaining differences in social interaction of children with a popular, an average, and a rejected sociometric status. Furthermore, we concentrate on two important aspects of interaction behavior, namely social instrumental actions and emotional expressions. The literature points to significant differences in these behavioral aspects between children of different sociometric statuses (Rubin, Bukowski, & Parker, 1998; Black & Logan, 1995, Eisenberg, Fabes, Guthrie, & Reiser, 2001; Hubbard, 2001; Denham, McKinley, Couloud, & Hold, 1990: Coie, Dodge, & Kupersmidt, 1990; Cillessen & Mayeux, 2004).

Goal and structure of this chapter

The goal of the model, namely to explain differences in social interaction, can only be accomplished, if the model has been empirically validated. Therefore, the primary aim of this chapter is to report about the empirical validation of our dynamic sys-
tems model of dyadic interaction of two children of different sociometric statuses. Two research questions follow from this aim, namely; first, what are the tools and steps for empirically validating this and comparable models. Second, how good is the model in representing the interaction process in reality? The answer to these two questions is conditional to answering an underlying question: What does the model explain about differences in short-term social interaction of children, especially in relation with children’s sociometric statuses, which is a property that tends to change over the longer term?

The structure of the chapter is as follows. This introductory section will be continued with a general description of the model and the way in which the model attempts to explain differences in social interaction. After this, we present criteria for an adequate model. For clarity reasons, method and results section are presented together. We subsequently present the empirical data with which the model is validated, and the statistical procedure and specific fitting methods that we used. In the discussion, we will - among others - reflect on the implication of our findings for using this type of simulation models in the field of research on social development.

**A dynamic systems model of the process of interaction**

The model is a combination of a dynamic systems model and an agent model in which two agents (two children) interact with each other. Examples of other agent models are described in Gilbert & Troitzsch (1999); Jager (2000), Jager, Popping, & van de Sande (2001); Kohler & Gumerman (2000); Staller & Petta (2001), Conte & Paolucci (2001), and Conte, Hegselmann, & Terna, (1997). For examples of other dynamic systems models of dyadic and group interaction, we refer to Buder (1991), Gottman, Guralnick, Wilson, Swanson, & Murray (1997); Felmlee & Greenberg (1999); Felmlee (2004); Warner (1992); Olthof & Kunnen (2000); van Geert (1994), and Steenbeek & van Geert (2005b). In addition, for more general information about the use of dynamic systems models in social developmental psychology in the broad sense of the word, we refer to van Geert (1994, 2003), Valsiner (1998); Granic & Hollenstein (2003). Fogel, (1993, 2001); Lewis (1995; 2000, 2002a); Lewis & Granic (2001); Dishion, Bullock, and Granic (2002).

The simulation model is based on a general theoretical model. For more explanation about the model’s general theoretical assumptions and the assumptions guiding the choice of the input parameter groups, we refer to Steenbeek & Van Geert...
Chapter 4

(2005a; Chapter 3). The mathematical elaboration of the model, in the form of annotated program code, is explained in more detail at the website www.gmw.rug.nl/~model/.

Theoretical assumptions that refer to a general theory of action and social interaction

The theoretical model comprises four assumptions about action in general and social interaction in particular. The first assumption is that action is goal- or intention-driven (Austin & Vancouver, 1996). Goals or intentions are often implicit, automatic or largely unconsciously operating. In addition, goals emerge under the control of the context, in the sense that “the environment itself activates and puts the goal into motion” (Bargh & Chartrand, 1999, p. 468; see also Austin & Vancouver, 1996; Ferguson & Bargh, 2004; Bargh & Ferguson, 2000).

The second general assumption is that behavior is aimed at achieving a maximal possible pleasure in a particular action situation, given the constraints and possibilities of that situation. This assumption refers to the idea that goals represent interests, or, as Frijda (1986) calls it, concerns. Concerns imply that organisms automatically evaluate situations in function of their goals, i.e. as either being good or bad 1. These evaluations are also called “appraisals”, which take the form of emotions (Scherer, 1999; Frijda, 1986). Appraisals ultimately refer to a general underlying evaluative dimension of pleasure versus displeasure, which has important potential for action (Cabanac, 1992; Johnston, 2003). Maximum pleasure can be achieved by means of withdrawal from unpleasurable situations and, alternatively, by approaching or constituting pleasure-enhancing situations (Austin & Vancouver, 1996; Roseman & Evdokas, 2004).

The third assumption is the principle of interaction, which entails three aspects. The first aspect points to interaction as an important goal itself. Interaction can be defined as “dyadic behavior in which the participant’s actions are interdependent such that each actor’s behavior is both a response to and stimulus for the other participant’s behavior” (Rubin, Bukowski, & Parker, 1998, p. 624). The second aspect of this principle follows from this definition, namely that interaction contains both a dyadic and an individual component. The dyadic component is the level of “reciprocity”

1 Note that, although many if not most concerns are largely “automatic”, some concerns are of course conscious and deliberate such as a person’s wish to buy a particular book or CD.
between interaction partners, which can be described as the “balance of all behaviors of both partners in an interaction that is built around a shared topic or mutual goal” (Linam, 1998, p. 4749). The individual aspect contains each child’s other-directed-behavior, which we define as any form of action addressed towards the other person. It refers, first, to *instrumental actions*, such as actual playing together, approaching the other child verbally or non-verbally, initiating a common action or trying to do so. Second, it refers to *positive emotional expressions*, such as smiles and laughter. Positive emotional expressions constitute the third aspect of the principle of interaction, namely they form an important constituent of approach-directed interaction. Note that we refer to positive emotional expressions in the first place, because they have a clear approach function. Note further that not all the other-directed behaviors that an individual child aims at interacting with another person, necessarily results in true interaction. The other person can avoid contact, refuse to respond or simply not notice the other’s person’s communication attempt.

The fourth assumption entails that behavior, including emotion, has a non-intentional component with two aspects. The first aspect refers to contagiousness, i.e. people tend to imitate each other’s behaviors. This concerns not only the behaviors per se but also more “abstract” and general aspects of that behavior. This non-intentional component is described in social learning and modeling literature (Bandura & Walters, 1977); entailing phenomena such as behavior contagion (Levy & Nail, 1993), and mood contagion (Neumann & Strack, 2000). It is also described in biologically and evolutionary inspired research about the direct coupling between perception and action (Preston & de Waal, 2002), and in neurophysiological literature, referring to the existence of mirror neurons (Rizolatti & Craighero, 2004). The second aspect refers to continuity of behavior, i.e. a certain automatic tendency to remain in one’s action mode until this tendency is overruled by an intentional drive towards a different action. A comparable concept is that of behavioral momentum (Nevin, 1988).

**Theoretical assumptions that refer to differential aspects of social interaction; the issue of sociometric status**

We assume that two important differential components contribute to differences in social interaction, namely social effectiveness and social power. Social effectiveness, or social competence (Rose-Krasnor, 1997), can be defined as “the ability to achieve personal goals in social interaction while simultaneously maintaining positive relationships with others over time and across situations” (Rubin, Bukowski, &
Research shows that popular children are more socially effective in interaction than rejected children (Simeo-Munson, 2000; Eisenberg & Fabes, 1995; Hazen, Black, & Fleming-Johnson, 1984; Hazen, Black, 1990; Hubbard, 2001; Asarnow & Callan, 1985; Coie, Dodge, & Kupersmidt, 1990; Black & Logan, 1995). Thus, the social effectiveness component of our model entails that higher-status (i.e. popular) children are more socially effective than lower-status (rejected) children. This implies that the popular child can make the play partner respond more often, i.e. his initiatives in involvement-behavior will be more often reciprocated by involvement-behavior of the play partner.

Social power is defined by Lewin (cited in Bruins, 1999, p. 8) as “the possibility of inducing force on someone else, or more formally, as the maximum force person A can induce on person B divided by the maximum resistance that B can offer”. One type of power is referent power, which is related to being the “best liked member of the group”, i.e. to having a popular status (Raven, 1992). In a dyadic situation, this difference in power will influence the interaction between the participants (Snyder & Kiviniemi, 2001). According to Copeland (1994), low power individuals (for instance, a child with a rejected or average status who interacts with a child with a popular status) are particularly motivated “to get along with the other person” (see also Dépret & Fiske, 1999). Thus, the social power component of our model entails that play partners of higher status (i.e. popular) children will have a higher motivation to get along with this high(er)-status child than play partners of lower-status (i.e. rejected) children. This higher motivation implies that the play partner will show more involvement-behavior, i.e. more instrumental actions directed towards the other child and more positive expressions (Snyder & Kiviniemi, 2001).

Theoretical principles of the dynamic systems approach as an integrative framework

According to dynamic systems theory, reality can be understood as consisting of dynamical systems that produce “especially complex, nonlinear behavior over time” (Thelen & Smith, 1998, p. 575). These dynamical systems must be studied as processes over time, instead of focusing on relations between variables at one moment in time. In this view, interaction is seen as an iterative, short-term dynamic pro-

---

2. Social competence and social effectiveness are exchangeable terms. We prefer to use the term ‘effectiveness’, because we emphasize the effectiveness-aspect of social competence in interaction, i.e. being able to achieve more by doing less (see Steenbeek & van Geert, 2005b).
cess, which emerges on the basis of mutual relationships between components of the interaction process. These components are –among others - the interaction partners’ goal-directedness, their appraisals, and their social effectiveness and social power (which are described in the previous theoretical sections). Dynamic principles, such as recursiveness, non-linear change, self-organization, the influence of chance, and the existence of attractor states, play a central role in this process\(^3\). For our model, one dynamic principle of interaction is of particular importance, namely the principle of “soft-assembly” (Thelen & Smith, 1994). In interaction, soft-assembly can be defined as the process through which the child’s interaction potentials, resulting from former interaction processes, are actualized as a specific, temporary level of interaction abilities of the child, in the course of this particular interaction (process) itself. The building blocks for this soft-assembly process are formed by factors that consist of the child’s own and the play partner’s properties and behaviors, which are for the most part incorporated in the model’s theoretical assumptions. These factors dynamically intertwine, i.e. at any moment in time, one child affects the other and the other child affects the first, thus creating the conditions under which both participants will operate during the next moment in time.

Practically speaking, there is no other way to actually understand this intertwining of multiple factors than by modeling it in the form of a dynamic, iterative process (Christiansen & Kirby, 2003). This brings us to discussing the basic components of the simulation model in the next paragraph.

**Implementation of variables in the simulation model**

*Input parameter groups*

The theoretical assumptions described above are reflected in the model by a number of input parameters. These input parameters are comprised in five input parameter groups. Table 4.1 gives an overview of these five input parameter groups.

---

\(^3\) For a thorough explanation of these principles, and how they are applied in the study of social emotional development, we refer to Thelen & Smith (1994); Fogel (1993, 2001); Lewis & Granic (2001); Newtson (1993); and Gottman, Murray, Swanson, Tyson, & Swanson (2002); van Geert (1994, 2003); van Geert & Steenbeek (2005).
Chapter 4

Table 4.1. An overview of the groups of input parameters, the specific parameters that are distinguished, and what can be adjusted

<table>
<thead>
<tr>
<th>Input parameter groups</th>
<th>Which parameters are distinguished</th>
<th>What can be adjusted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Concerns group</td>
<td>Involvement</td>
<td>Strength of the concerns in relation to each other</td>
</tr>
<tr>
<td></td>
<td>Autonomy</td>
<td></td>
</tr>
<tr>
<td>2 Realizability group</td>
<td>Influence of ‘playing together’</td>
<td>Strength of the influence of behavior on the realization of a concern</td>
</tr>
<tr>
<td></td>
<td>influence of ‘playing alone’</td>
<td></td>
</tr>
<tr>
<td>3 Expressiveness group</td>
<td>Positive expression</td>
<td>Ease with which emotional appraisal is translated into an emotional expression</td>
</tr>
<tr>
<td></td>
<td>Neutral expression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative expression</td>
<td></td>
</tr>
<tr>
<td>4 Preference group</td>
<td>Influence of positive expression</td>
<td>Strength of the influence of emotional expression on the preference of a concern</td>
</tr>
<tr>
<td></td>
<td>Influence of negative expression</td>
<td></td>
</tr>
<tr>
<td>5 Non-intentional behavior group</td>
<td>Continuity</td>
<td>Strength of the non-intentional principles in relation to each other</td>
</tr>
<tr>
<td></td>
<td>Symmetry</td>
<td></td>
</tr>
</tbody>
</table>

The first and most important parameter group relates to concerns. It represents one concern dimension that refers to a specific balance between the strength of the concern for Involvement, which is the tendency to direct one’s behavior towards the other person, and the strength of the concern for Autonomy, which is the tendency to perform a solitary action. Note that autonomy in this sense does not refer to autonomy as a personality characteristic. This parameter group specifies the child’s most pleasurable alternation between playing with the other child and playing on his own, during the current interaction. Note that the child’s realized balance between these concerns self-organises in a context-and-person specific way (Shaw, 2001; Gibbs & van Orden, 2003; Kappas, 2002).

The second group, or realizability group, represents the influence of behavior on the realization of concerns. For descriptive reasons, behavior is reduced to two main categories, ‘playing alone’ and ‘playing together’. Basically, in the model the child realizes his concern Involvement by showing the acting category ‘playing together’, and his concern Autonomy by showing ‘playing alone’. This group deter-

---

4 Note that by focusing on these two concerns in the concerns parameter group, we aim to concentrate on fundamental order parameters in the interaction process (for more information about the role of order parameters, see van Geert & Steenbeek, 2005). It does not mean that we deny the existence of other concerns in the process.
Validating the agent model

mines for instance to what extent the child feels an event of ‘playing together’ as pleasurable and thus contributes to realizing the child’s concern for Involvement.

The third parameter group, or expressiveness group, represents the strength of the relation between emotional appraisal and emotional expression. In the model, the expressions are simplified by reducing them to one quantitative dimension, ranging from positive, over neutral to negative. An example of what is being determined here is the ease with which a child shows a positive emotional expression, if the appraisal is positive.

The fourth parameter group, or preference group, represents the influence of emotional expression on the preference of concerns. The idea is that the child’s preferred level of a concern can change under the influence of his own and of his play partner’s emotional expression. For instance, the parameter specifies to what extent a positive expression that accompanies an event of ‘playing together’, makes playing together more desirable or pleasurable for this child. That is, to what extent does it enhance the child’s current preference for the concern Involvement?

The fifth - and last– parameter group refers to non-intentional basic principles of behavior. It represents one general behavioral dimension concerning the specific balance between the tendency to continue one’s own behavior (continuity), and the opposite tendency, namely to do what the other person is doing (symmetry, which depends on the contagiousness of the behavior of the other person).

During each model run, the parameter values are stochastically varied within preset limits.

Internal and external processes

The values of the input parameter groups co-determine an internal process that takes place within each child separately, at every moment (i.e. every time step in the model). This internal process can be described as follows: The process starts with the concerns of this child. The difference between the child’s preferred and realized value of each concern – at this particular moment - results in a certain strength of a drive. The concern with the highest drive will produce the behavior of the child. Via an appraisal function, the average of the drives will result in an emotional expression. For instance, if at this moment the concern Involvement produces the highest drive, the child will show the accompanying behavior ‘playing together’. In addition, the average of the values of the drives results in the child showing a positive expression.
The connection between two internal processes at two successive moments (for instance, time $t$ and time $t+1$) takes place via an external, i.e. context process that involves both children. In short, this external process can be described as follows. First, the behavior at time $t$ of a child has an influence on the realized level of the child’s and the play partner’s accompanying concern at time $t+1$. Second, the emotional expression at time $t$ of a child has an influence on the preferred level of the child’s and the play partner’s accompanying concern at time $t+1$. For instance, at time $t$ the child is ‘playing together’, and enjoys it, which shows in the big smile on his face. For time $t+1$, this implies that both the realized level of the concern Involvement and the preferred level of the concern Involvement are increased. On the basis of these new level values, it is now possible to calculate the behavior and emotional expression of both children at time $t+1$.

Note that in a complete simulation run of the model, these internal and external processes take place during and between each successive moment, i.e. each simulation step. For instance, in a play session of 7 minutes, in which each time step is one second, 420 such time steps – and thus internal and external processes - take place.

For a detailed description of these processes, we refer to Steenbeek and van Geert (2005a; Chapter 3).

**Output variables**

The output variables are the result of the internal and external process as described above and consist of each child’s involvement-behaviors (emotional expressions and instrumental actions), at every moment in the interaction process. That is, the simulated child yields an output from the first until the last moment of the play session, and so does the play partner. Thus, over the course of an interaction, a pattern of expressions and instrumental actions emerges for each child.

In order to be able to compare this output with empirical data, we transformed it into twenty-two operational variables that describe all potentially relevant aspects of involvement-behavior (see table 4.2).
Validating the agent model

Table 4.2: Operational variables, such as derived from the model output and derived from the empirical data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>child or play partner</td>
<td></td>
</tr>
<tr>
<td>Directedness</td>
<td>proportion of directed actions (‘playing together’) over the total of all actions (both ‘playing together’ and ‘playing alone’) of the child</td>
</tr>
<tr>
<td>Proportion shared directedness</td>
<td>proportion of directed actions of this child, accompanied by a directed action of the play partner</td>
</tr>
<tr>
<td>Positive expressions (^a)</td>
<td>proportion (percentage of time) of positive expressions over the total number of expressions (neutral, negative or positive expressions)</td>
</tr>
<tr>
<td>Intensity positive time (^a)</td>
<td>proportion of total average intensity of positive expressions</td>
</tr>
<tr>
<td>Intensity positive number</td>
<td>proportion of intensity of positive expressions divided by the amount of positive expressions</td>
</tr>
<tr>
<td>Proportion shared positive</td>
<td>proportion of positive expressions of this child, accompanied by a positive expression of the play partner</td>
</tr>
<tr>
<td>Negative expressions</td>
<td>proportion (percentage of time) of negative expressions over the total number of expressions (neutral, negative or positive expressions)</td>
</tr>
<tr>
<td>Intensity negative time</td>
<td>proportion of total average intensity of negative expressions</td>
</tr>
<tr>
<td>Intensity negative number</td>
<td>proportion of intensity of negative expressions divided by the amount of negative expressions</td>
</tr>
<tr>
<td>dyad</td>
<td></td>
</tr>
<tr>
<td>Coherence dyad</td>
<td>proportion of time that both children show directed actions (‘playing together’) of the total time of the play session</td>
</tr>
<tr>
<td>Shared positive expressions (^b)</td>
<td>proportion of shared positive expressions over the total number of expressions. This variable can be read as a measure for “coherence of positive expressions”</td>
</tr>
<tr>
<td>Shared negative expressions</td>
<td>proportion of shared negative expressions over the total number of expressions. This variable can be read as a measure for “coherence of negative expressions”</td>
</tr>
<tr>
<td>Contrast dyad</td>
<td>proportion of contrast in intensity of expressions of both children over the total time. The time that both children express a neutral expression is not included (coded as zero)</td>
</tr>
</tbody>
</table>

Notes. a. Concerning variables ‘positive expressions’ (1) and ‘intensity positive time’ (2): in (1) the intensity is not calculated, in (2) the proportion in relation to intensity.
b. Concerning variable ‘shared positive expression’ (of the dyad): A high level of shared positive expression does not necessarily imply a high level of positive expressions per se.
The operational variables refer to the duration of actions (5 variables) and the
duration or intensity of expressions (17 variables) of each child separately and of the
dyad. An example of an action-related variable is the child’s ‘directedness’; i.e. the
proportion of actions (‘playing together’) directed towards the other child, over the total
of all actions (both ‘playing together’ and ‘playing alone’; see first row of table 2; note
that this variable also exists for the play partner). If both children show directedness
simultaneously, for instance at time \( n \), the dyad is coded as ‘coherent’ at time \( n \). An
example of an expression-related variable is ‘positive expression’, which is the child’s
proportion of positive expressions over the total of all expressions (positive, neutral,
and negative). (See table 4.2, third row).

Subsequently, these operational variables can be averaged over a number of
model runs of a particular dyad group. These averaged operational variables can be
compared with our empirical data, since in the latter we also coded an ‘output’ per
second for each child, consisting of the child’s emotional expressions and instrumen-
tal actions. This ‘output’ was translated into the same twenty-two operational vari-
ables. We averaged them over the three groups of dyads distinguished in our empiri-
cal study.

How does the model explain differences and similarities in social interac-
tion?

We assume that a child with a specific sociometric status, in the context of
playing with another child with its own particular properties, is represented by means
of specific settings of the five input parameter groups. A further assumption is that
these settings depend on social effectiveness and social power; i.e. the differential
theoretical components of the model (see section ‘theoretical assumptions that refer
to differential aspects of social interaction’). Remember that these components refer
to the social effectiveness of the child itself, and to the influence of the social power of
the child’s play partner.

Table 4.3 shows the settings of the input parameter groups for children of dif-
ferent sociometric statuses. The table shows that we distinguish three types of dyads
of children. They consist of a child with either a rejected status playing with an ave-

---

5This is in accordance with the design of the broader study that we conducted in collaboration with the
University of Utrecht. The main goal of this collaborative project was to study differences in interaction of
popular and rejected children in a standardized play context, with dyads composed as described above
(see for instance Gerrits, 2004).
### Table 4.3. Settings of input parameters of children of different sociometric status in the context of playing with a play partner.

<table>
<thead>
<tr>
<th>Type of dyad</th>
<th>Rejected dyad</th>
<th>Average dyad</th>
<th>Popular dyad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status of child</td>
<td>Status of play partner</td>
<td>rejected</td>
<td>average</td>
</tr>
<tr>
<td>1 Concerns</td>
<td>I much stronger than A</td>
<td>I little bit stronger than A</td>
<td>I stronger than A</td>
</tr>
<tr>
<td>2 Influence of behaviour on realisation of concerns</td>
<td>I = average</td>
<td>I = high</td>
<td>I = average</td>
</tr>
<tr>
<td>3 Relation between emotional appraisal and emotional expression</td>
<td>positive = moderate</td>
<td>negative = difficult</td>
<td>positive = moderate</td>
</tr>
<tr>
<td>4 Influence of emotional expression to preference of concerns</td>
<td>positive = big</td>
<td>negative = average</td>
<td>positive = average</td>
</tr>
<tr>
<td>5 Non-intentional principles of behaviour</td>
<td>continuity = average</td>
<td>symmetry = high</td>
<td>continuity = average</td>
</tr>
</tbody>
</table>

**Notes.**

a. The table can be read as follows: a child with a rejected status that plays with a play partner with an average status (rejected dyad; upper left of the table) has as setting for the first input parameter group Concerns that his / her Concern Involvement is much stronger than his / her Concern Autonomy, which is expressed as 'I much stronger than A'.

b. I = Concern Involvement, A = Concern Autonomy. The average dyad has standard settings (average, moderate). Everything that differs from these standard settings is printed in italics (high, low, difficult, big).

c. Information about the corresponding ranges of numerical values of these settings of input parameter groups for children of different statuses in the context of...
rage play partner (the “rejected” dyad); a child with a popular status child playing with an average play partner (the “popular” dyad), or a child with an average status playing with a play partner with an average status (the “average” dyad). By means of this table, we will illustrate the rationale behind those settings.

First of all, we focus on the settings for the average child, playing with an average play partner. Our starting point is that this “average” dyad has default values for the five input parameter groups. These default values have been obtained by estimating the best possible set of model parameters, based on a qualitative comparison with the empirical data from “average” dyads. Note that for reasons of simplicity these quantitative parameter settings are specified in qualitative terms (such as ‘average’, ‘stronger than’ etc.; the accompanying numerical values can be found in the parameter worksheet of the simulation model which is available in the form of an excel-file at website www.gmw.rug.nl/~model/, see also Appendix B).

The children in the “average” dyad both have an average status, thus they are average in both their child specific (internal) properties and their context specific properties. It means that the settings of both children refer to an average influence of the social effectiveness component and the social power component. Table 3 shows that for instance for the concerns group of input parameters, the default setting is that Involvement is stronger than Autonomy. In addition, both the behavior ‘playing together’ and the behavior ‘playing alone’ have an average influence on the realization value of the corresponding concern, respectively the concern Involvement and the concern Autonomy (realizability group of input parameters).

For a discussion and mathematical description of the influence of the social effectiveness and social power components for the “rejected” and “popular” dyads, we refer to the more extensive description on the website www.gmw.rug.nl/~model/. It can be shown on mathematical grounds that as far as the parameter values are concerned, a combination of these influences leads to a division into two groups. The first group consists of the rejected child in the “rejected” dyad and the average child in the “popular” dyad. These children have the same deviation from the default settings, namely one in which the concern Involvement is much stronger than the concern Autonomy, the realizability of Autonomy is high, the likelihood of negative expressions is low, the preference effect of both positive and negative expressions is big, and the tendency to symmetry is high. This means that, for instance, the rejected child has a high concern for playing together (Involvement). He is easily satisfied with ‘playing alone’, and tends to suppress negative emotional expressions. In addition, the play
Validating the agent model

partner’s positive expressions during ‘playing together’ will strongly increase the child’s preference for ‘playing together’. Finally, his tendency to imitate the behaviors of the play partner is strong.

The second group consists of the average child in the ‘rejected’ dyad and the popular child in the ‘popular’ dyad. These children have the following non-default settings: their concern Involvement is a little bit stronger than the concern Autonomy, the realizability of Involvement is high, and the tendency to symmetry is low.

Note again that these settings correspond with preset quantitative values.

What are the criteria for a good model and a good empirical fit?

An important general criterion for a good model is that it “provides convincing answers to the questions we put to it” (Casti, 1997; p. 25). In addition, Gottman (Gottman et al., 2002; p. 67) states that a good model “gets people to ask questions, especially ones that help generate ideas for new experiments within the science, thus keeping the interplay going”. Put more concretely, it means that, first, the model must be based on valid theory (-ies), with valid definitions of behavior; and that, second, the model must technically “behave well”, for instance not showing chaotic patterns, if in reality the represented system does not either. Thus, in terms of sensitivity of the model, it must neither be too sensitive nor too insensitive to changes in values of input parameter groups. Concerning the fit with the empirical data, the model must show sufficient similarity in its output compared to results of empirical observations of the process (see Balci, 1997; Van Dijkum, DeTombe, & van Kuijk, 1999; Gilbert & Troitzsch, 1999). For our model, it means that interaction properties for children of different sociometric statuses must be generated, comparable with those in empirical findings. In order to do so, the choice of parameters must be theoretically adequate, but also must be the best possible choice. The question is of course what is ‘sufficient similarity’ and ‘comparable output’. This question will be answered in the method section.

A necessary final remark is that our model deals with a process, thus it would be obvious to focus the fit criteria on process characteristics, especially those concerning patterns of the output variables over the course of an interaction. However, an important first question is whether the model gives a good representation of general characteristics of the sample, namely of the averages and distributions of the different status groups. On the other hand, it makes little sense to fit group characteris-
tics if the patterns generated by the model were to deviate considerably from the empirical patterns. We checked for the existence of such deviations by means of a qualitative procedure, which showed that the model yields patterns with sufficient qualitative similarity with empirical data, especially concerning clusters and peaks of specific behaviors and emotional expressions (Rhemrev & de Haan, 2003; van der Singel, 2003; Steenbeek & van Geert, 2002a; van Geert & Steenbeek, 2005).

Recapitulation of research questions

Recapitulating, the goal of this chapter is to validate our interaction model with empirical data. Our research questions are twofold. First, what are the tools and steps by means of which our model can be validated with empirical data? This question implies not only whether the choice of parameters is adequate, but also whether the choice of parameters is the best possible. Second, how good is the resulting fit? Finally, if these two questions have been answered sufficiently, the question can be answered: What does the model teach us about differences in social interaction of six-, and seven year-old children of different sociometric status?

METHOD AND RESULTS

Empirical data

In this section, we will present the main lines of our empirical study. For a more elaborate discussion of these empirical data, we refer to Chapter 2 (Steenbeek and van Geert, 2005b; Chapter 2).

Participants: 24 dyads of grade 1 pupils with mean age of 6.5 years participated. They were selected on the basis of their sociometric status. Each dyad consisted of two same-sex children. Three types of dyads were formed, comparable with the dyads distinguished in the model. The first consisted of a child with a rejected status, coupled with a neutral play partner with an average status (“rejected” dyads). The second group consisted of a child with a popular status, and an average play partner (“popular” dyads); the third of a child with an average status, and an average play partner (“average” dyads).

Procedure: First, the sociometric status of the participants was determined on the basis of repeated measures of a rating test (Asher, Singleton, Tinsley, & Hymel, 1979), which was analyzed with the computer program SS-rat (Maassen, Akkermans,
Validating the agent model

& van der Linden, 1996; the stability of the procedure is discussed in Maassen, Steenbeek, & van Geert, 2004; Steenbeek & van Geert, 2005c). Second, the dyads were videotaped during a ten-minute play session, which took place in a separate room in the school. The only instruction was to play together with four groups of toys that were placed on the table. After giving the instruction, the researcher left the room, leaving the children alone with the toys and the camera.

The 24 dyads were videotaped three times, with intervals of approximately one and a half month. In principle the second and third round were selected for coding. Due to practical limitations, only 17 dyads were coded twice and 7 dyads were coded once. This resulted in a total of 41 coded interactions (“rejected” dyads; n = 13; “average” dyads; n = 14; “popular” dyads; n = 14) 6.

Coding and variables: The recordings were coded with the computerized system Observer 4.0 pro (Noldus Information technology, 1999). To determine the inter-observer reliability between the observers, we used a nonparametric permutation test (see section ‘random permutation analysis’ for more information about permutation techniques; see also van Geert & van Dijk, 2003). The reliability was determined in advance, and can be considered good in terms of percentage agreement (.8 for coherence, p = .01, .81 for expressions, p = .01).

Two variables were coded: emotional expressions and instrumental actions of each child separately. Changes in expressiveness and action of each videotaped child separately were coded with a precision of 1/10 of a second (event sampling). The variable emotional expression was coded on a scale ranging from very negative (-4) to very positive (+5), representing the intensity of the expression. Categories -4 to -2 represented negative expressions, -1 to +1 neutral expressions, and +2 to +5 positive expressions. The variable action was coded with the help of three overt variables: verbal turn, nonverbal turn, and focus. On the basis of these partial variables, a child’s current behavior is coded as directed action (‘playing together’) or not (‘playing alone’). If the child displays neither a verbal turn, nor a nonverbal turn, nor a focus (towards the play partner or the mutual play activity), the child is supposed to display non-directed actions (‘playing alone’). Otherwise the child is coded as displaying directed actions (‘playing together’). Remember that directed action also entails

6 Since we use permutation methods there is no problem with the fact that in a number of cases the same dyads occur twice in the sample. There is no assumption of independent measures. The small sample sizes are explained by the fact that popular and rejected children form a small minority of the total number of children in the class.
attempts towards involving the other child in the interaction. If both the child and the play partner show mutually responsive directed actions, the behavior is coded as “coherence dyad”, which is the only action variable on the dyad level.

In order to derive as much information as possible from the variables expression and action, they were translated in operational variables, both on the individual and on the dyadic level. They are exactly the same as those described in table 2 (see section ‘output variables’).

Results: What follows is a short summary of the findings presented in Steenbeek and van Geert (2005b). First of all, the total number of twenty-two variables was reduced to a subset of ten core variables. Over the total pattern of this selection, we found significant differences and similarities between “rejected” dyads and “popular” dyads ($\chi = 7.6, p = 0.006; \chi^2 = 8.6, p = 0.001$).

Secondly, by examining each of the twenty-two variables separately, we found statistically significant differences between “rejected” dyads and “popular” dyads in seven variables. In six of these variables, the “rejected” dyads scored significantly higher than the “popular” dyads, namely in the child’s ‘directedness’, ‘positive expressions’, and ‘intensity positive time’; in the play partner’s ‘proportion shared directedness’ and ‘intensity negative number’; and finally in the dyad’s ‘coherence’. Note that the findings concerning positive expressions and directedness seem counterintuitive, because popular children are often described as showing more positive expressions and directedness.

Only in one variable of this set of seven, namely the child’s ‘proportion shared positive’, the “popular” dyads scored significantly higher than the “rejected” dyads. This finding is consistent with the literature, which points to the existence of more ‘mutuality’ in interactions in which a popular child is involved.

We did not find significant differences between “rejected” dyads and “popular” dyads in ten variables, namely the child’s ‘proportion shared directedness’, ‘intensity positive number’, and ‘intensity negative number’; the play partner’s ‘directedness’, ‘positive expressions’, ‘intensity positive time’, ‘intensity positive number’ and ‘proportion shared positive’; and the dyad’s ‘shared negative expressions’, and ‘contrast child–play partner’. This means that contrary to what has been reported in

---

7. Results over the “average” dyads are not reported, because analyses reveal that in all variables but one the mean score of the “average” dyads lie in between the mean scores of the “popular” dyads and the “rejected” dyads. The one exception is the variable ‘intensity negative number’ of the child, in which the difference is not statistically significant.
the literature about comparable variables, we did not find differences for “rejected” and “popular” dyads. For instance, we did not find that the play partner of a popular child directs his actions more often to the popular child than the play partner of a rejected child directs his actions to the rejected child. Finally, five variables showed a trend (with \( p \)-values between 0.1 and 0.3). These variables were the child’s ‘negative expression’ and ‘intensity negative time’; the play partner’s ‘negative expression’ and ‘intensity negative time’; and the dyad’s ‘shared positive expressions’. This again is not consistent with what has been found in literature earlier. The trend found in our data is that popular children show more ‘negative expressions’ than rejected children; and popular dyads show less ‘shared positive expressions’ than rejected dyads, whereas literature reports more negative expressions in rejected children, and more positive expressions in popular children. In addition, the trend found in our data is that play partners of popular children show less ‘negative expressions’ than play partners of rejected children.

The discrepancy between our findings and earlier findings can be explained as follows. First, we used a process model to make predictions, which can easily differ from those that depart from child-specific factors found by averaging over many interactions. Second, we used a specific play situation in our research setting, namely one in which the children were explicitly asked to play with the other child, and where adults monitored them. Note that this situation is different from a free play situation in the daily life of children. The fact that we found rejected children to behave in this way in this particular situation does not not exclude another fact, described in literature discussed earlier, namely that their overall daily experience might be one of predominantly negative emotions and little interaction. A challenge for our process model is to expand our predictions to other interaction situations, and try to explain why there are infrequent and often negatively loaded interactions for rejected children.

**Procedure**

**Generation of model output**

The output of the model is obtained by running the model 5000 times for each set of input parameters that represents a specific type of dyad. All averages and distributions of the operational variables resulting from the model runs of different types of dyads are significantly different. Note that every run also contained a random influence, contributing to the typical distributions of the end results of each status group.
Random permutation analysis

The statistical problem is that a total of 5000 model runs of dyads with a specific status is compared with 13, 14, and 14 empirical dyads with that particular status, respectively. To solve the problem of the major difference in size between empirical and model samples, we use random permutation analysis, which are in particular convenient for small and unbalanced datasets. This technique is highly flexible, and can test explicitly formulated null hypotheses (Good, 1999; Manley, 1997; Todman & Dugard, 2001). The basic principle is that it estimates the probability that an observed result is caused by chance alone. This is done by simulating that probability, i.e. by drawing a very large number of accidental samples (e.g. of scores of popular and rejected children), and then counting the number of times that the observed phenomenon (or an even ‘stronger’ one) occurs in the accidental samples. To reduce the estimation error, random samples were drawn 5000 times.

A broader theoretical justification of the procedures used in this chapter is as follows. In the present kind of research, basically two pictures of reality are generated. One is a picture based on observations (the empirical study), the other is based on the simulation model. The question is: to what extent do the two pictures resemble each other? A good resemblance does not imply that the pictures look alike to the level of the smallest details, but there should be sufficient similarity so that an unbiased viewer can recognize one picture as a representation of the other. The hypothesis is that this similarity is meaningful. The nullhypothesis is that this similarity could have been produced by chance alone.

However, how is chance defined? The simplest form of chance consists of randomly shuffling the ‘output’ of the model. This ‘output’ consists of expression and action variables for the different dyad groups, as expressed in averages and in distributions. This form of similarity testing will be conducted with the pattern analysis of averages and distributions (see section ‘fitting averages’ and ‘fitting distributions’). Basically, the assumption behind the null hypothesis is that the output of the model can not be distinguished from a purely random coupling between output variables and types of dyads.

A second form of chance, which is related to a stricter form of null hypothesis testing, is represented by randomly shuffling the input of the model. This input consists of the parameter settings. This form of randomization more explicitly modifies
Validating the agent model

the basic assumptions of the model, which are represented in the form of specific sets of values that are coupled with the distinct types of dyads. This type of null hypothesis testing will be applied with the distribution analysis over distinct variables, mainly because in this particular case the simpler form of null hypothesis testing is not applicable (see Uebersax, 2005; van Geert & van Dijk, 2003). Basically, the assumption behind the null hypothesis model is that the assignment of specific parameter values to specific types of dyads is coincidental, and thus that any such random assignment will result in empirical fits that are as good or as bad as those found with the real model.

Operationalization of the research questions - an overview of the fitting methods

We answer the question by means of which tools and steps the empirical data can be fitted to the model by comparing the operational variables in three manners. First, we compare the average of each operational variable over 5000 model runs - per type of dyad - with the empirical averages of 13, 14, and 14 dyads, respectively. Second, we compare the distributions of each operational variable over 5000 model runs with the empirical distributions. Third, a sensitivity analysis is conducted, in which we check whether the model is too sensitive or too insensitive for changes in parameter values. This three-fold method can be justified as follows. First, in addition to using averages we also fit distributions, which provide information about an important aspect of the data, namely the nature of the differences between individual members of the same dyad groups. Second, by performing sensitivity analysis, we attempt to find ranges of adequate values of input parameter groups, instead of confining ourselves to estimating single values. It is likely that such ranges provide a more valid representation of particular dyad groups than single central values.

With the help of these three fitting methods, we answer the second question, namely how well the model fits the data. In answering the first aspect of this second question, namely to what extent do the chosen parameter values fit the data, we use our comparison of averages and distributions of each variable. The second aspect is whether the chosen parameter settings are the best possible choice of parameter settings. This aspect can be answered by means of sensitivity analysis. (Note that sensitivity analysis also contributes to answering the first aspect of this question).

Given the complexity of the method, the following section combines the presentation of the method and results.
**Fitting averages**

The question “does the model give a good representation of the group averages?” is divided into two specific questions. The first relates to a comparison between model and data for the pattern of all variables together, over the three types of dyads. The second involves a comparison between the simulated and the empirical averages of each variable separately for the distinct types of dyads.

*The pattern of all variables, over all status groups*

The specific question is: does the model give a good representation of the pattern of empirical averages, namely of the combination of all variables over all status groups. In other words: does the systematic, model-dependent, combination of values with variables in a particular dyad group, bear a better resemblance with the empirical pattern than a random combination of values with variables does? The null hypothesis is that the simulated pattern does not give a better representation of the pattern of empirical averages than a random one.

The statistical method: First, we rescaled the values of the simulated and empirical averages of all twenty-two variables by dividing them by their maximum value. By doing so, it is possible to compare values of distinct variables. Second, the distance between the simulated and the corresponding empirical value is calculated, of each separate variable, of each separate status group. The distance is expressed in the form of a chi, i.e. the absolute value of the difference between the simulated and the observed value of the variable. Third, these distances are summed over all variables, over all status groups. The resulting value of this criterion is compared with chance by means of a random permutation method. By doing so, we calculated the probability that a random model yields a sum of chi’s that is as small as or smaller than the actual simulation model does.

Results: the criterion (the sum of chi’s) based on the actual simulation model is significantly smaller than can be expected on the basis of chance ($p = .01$). This means that the simulated pattern gives a better than chance resemblance with the pattern of empirical averages.

*Per variable, per status group*

---

8. We also calculated an alternative measure of distance, namely chi square, and a stricter rescaling version, in which $x - \min x / (\max x - \min x)$. These alternative versions also produce significant results; in the chi-square ($p = .05$), and in the stricter version of rescaling ($p < .01$).
The specific question is: does the model give a good representation of the empirical averages, for each variable separately? This question can be operationalized in two ways: first: does the simulated average falls within the range of the empirical confidence interval? Second: is the order of ranking of the simulated averages of the three groups the same as the empirical one? Checking the order of ranking makes sense only for variables that show empirical differences between ‘popular’ dyads and ‘rejected’ dyads (see section ‘empirical data’).

*The statistical method:* The 95% confidence interval of the empirical values is determined by resampling the data 1000 times, via a bootstrap-procedure (Efron, 1988). This way, a range of possible values for each variable results that could have been obtained empirically, if the sampling had, accidentally, been different.

*Results:* Figure 4.1 provides an example of the position of the simulated averages compared to that of the empirical confidence range, for the variable ‘*intensity positive expressions child*’.

The simulated averages of the three types of dyads lie within the confidence

![intensity of expression over time in the child](image)

*Figure 4.1.* An example of the position of the simulated averages compared to those of the empirical confidence range, for the variable ‘*intensity positive expression child*’.
intervals of the empirical averages. In addition, in the three dyad groups, the order of ranking of the simulated averages is the same as the empirical ranking.

Table 4.4 shows the fit quality of all variables for the three types of dyads. First, we will discuss the fit quality as expressed in the confidence interval. The fit quality refers to the distance between the variable generated by the model and the confidence interval of the empirical variable. It is defined as follows: Let V be the value produced by the model, L the lower boundary, and U the upper boundary of the empirical confidence interval. The fit quality is calculated as follows: \(1-(|V - L| + |V - U| - |U - L|)\). A fit quality of 1 means that the simulated average falls within the empirical confidence interval, smaller values reflect increasing distances.

### Table 4.4. Results of fitting procedure over averages; per variable, per status group

<table>
<thead>
<tr>
<th></th>
<th>confidence interval</th>
<th>order of ranking</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rejected</td>
<td>average</td>
<td>popular</td>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.92)</td>
<td>(1)</td>
<td>(0.84)</td>
<td>(0.92)</td>
<td>(0.04)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Proportion shared directedness</td>
<td>(0.84)</td>
<td>(0.96)</td>
<td>(0.75)</td>
<td>(0.85)</td>
<td>(0.07)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Positive expressions</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(0)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Intensity positive time</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(0.01)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Intensity positive number</td>
<td>(1)</td>
<td>(0.65)</td>
<td>(0.61)</td>
<td>(0.75)</td>
<td>(0.04)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Proportion shared positive</td>
<td>(0.87)</td>
<td>(1)</td>
<td>(0.73)</td>
<td>(0.87)</td>
<td>(0.1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Negative expressions</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Intensity negative time</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Intensity negative number</td>
<td>(1)</td>
<td>(1)</td>
<td>(0.59)</td>
<td>(0.86)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
</tbody>
</table>

Notes. a. The fit quality can be interpreted as follows: For the confidence interval, a fit quality of 1 means that the simulated average falls within the empirical confidence interval, smaller values reflect increasing distances. For the order of ranking, a fit quality of 1 means that the rank order of the simulated average is the same as the rank order of the empirical average; 0 means that these rank orders are not the same.

b. The seven variables that yielded significant differences between status groups in the empirical study were printed in **italics.**
For example, the fit quality for the variable child’s ‘directedness’ is 0.92, 1, and 0.84, for the three dyad groups, with a $p$-value of 0.04. All $p$-values smaller than .05 are considered good. If the $p$-value is < .001, the fit is considered excellent, between 0.05 and 0.1 moderate, and if $p$ is greater than 0.1, the fit is considered poor. Overall, the fit quality per variable is excellent for 10 variables, good for 6 variables, moderate for 4 variables, and poor for 2 variables.

The fit quality per type of dyad is specified in the last two rows of the table. The three $p$-values are smaller than 0.01, which implies that they are significantly better than chance. The rejected and average dyads fit equally well (fit qualities of 0.95 and 0.97). The popular dyad fits less well (fit quality of 0.85), and this difference in fit quality is statistically significant ($p < 0.01$).

Second, we will discuss the fit quality as expressed in the order of ranking. Note that we report only the twelve variables for which rank order is meaningful, namely those that are empirically different. (The cut-off value was $p < .3$; seven of these twelve had a $p$-value smaller than .05). Note further that since calculating $p$-values for separate variables makes very little sense (the number of possible permutations is too small), we only report the $p$-values for the three types of dyads. For the ‘rejected’ dyads and the ‘popular’ dyads, the fit is moderate ($p = 0.06$). For the average dyads, the fit is poor ($p = 0.17$). Over all dyads, the fit is excellent ($p < 0.01$).

**Fitting distributions**

The question is: do the simulated model distributions give a good representation of the observed (empirical) distributions of all variables? In accordance with the discussion of the averages, we make a distinction between the fit over the total pattern of distributions, and the fit over the variables separately.

**The pattern of all variables, over all status groups**

The specific question is: does the model give a good representation of the total pattern of distributions of all variables, over the three status groups? This question is answered with the help of two criteria, namely the position of the coordinates of the peak in the distribution, and the histogram of the distribution itself. Just as in the analyses of the averages (previous paragraph), we assume that the model yields a specific pattern of histograms. This pattern amounts to a specific ordering of elements, in this case histograms. The null hypothesis is that the ordering of these ele-
ments is accidental; i.e. the random model assumes that there exists no systematic association between a particular histogram and a particular variable.

**Criterion: coordinates of the peak**

The question is operationalized as follows: do the coordinates of the peak of the simulated histograms resemble the empirical coordinates more than can be expected on the basis of chance? Coordinates of the peak are both the position on the y-axis (the maximal frequency) and the position on the x-axis (the value that occurs most frequently).

The **statistical method**: First, histograms are determined for the empirical data, the model data, and the random model data (recall that the random model consists of a series of random couplings of distributions to variables). Second, for each variable (in the model data, the empirical data, and the random model data), the coordinates of the peak are determined. Third, for each variable we calculate the distance between the coordinates of the peak in the model distribution and the coordinates of the peak in the empirical distribution. In addition, for each variable we calculate the distance between the coordinates of the peak in the *random* model and the coordinates of the peak in the empirical distribution. Fourth, we average the distance over all variables for the two coordinates separately, and then average over the two coordinates (*chi criterion*). Finally, this criterion is compared with averages obtained on the basis of the random model.

**Results**: The distance between the average of the coordinates of the peak in the model distribution and the average of the coordinates of the peak in the empirical distribution is significantly smaller (*p* < .001) than can be expected on the basis of chance.

**Criterion: histograms**

The question is whether the simulated histograms of the distribution resemble the histograms of the empirical distribution more than can be expected on the basis of chance. The null hypothesis is that the histogram of variable A could just as well belong to variable B or C; i.e. if one gets an arbitrary sample of histograms, one also would obtain a good fit with the histograms of the empirical distribution.

The **statistical method** can be compared with the method described for the coordinates of the peak. The only difference is that we now focus on values for each distinct bin in the histogram.

**Results**: The distance between the histograms of the model distribution and those of the empirical distribution, for all variables for all status groups, is significantly
Validating the agent model

smaller ($\rho < .001$) than the corresponding distance between the histograms of the random model distribution and those of the empirical distribution.

**Per variable, per status group: testing a model with an arbitrary parameter range**

The specific question is: does the model give a good representation of the distributions of all variables separately, over all status groups separately? Remember that the comparison between the model and a null hypothesis model (which specifies chance) differs from preceding analyses. The current null hypothesis model randomizes the values of the *input* parameters (see section ‘random permutation analysis’).

Each distribution is represented by a histogram, based on 11 bins (one bin for the 0-value, one for values between 0 and 0.1, another for values between 0.1 and 0.2, etc; most variables range from 0 – 1; 5 variables referring to emotional expression have a considerably broader range). For each variable, there are two distributions. One is based on the proportion of frequencies of the observed values; the other is based on the proportion of frequencies from 5000 runs of the model (e.g. the 13 values of “coherence” of the “rejected” dyad, and the set of 5000 ‘coherence’ outputs from the model). The hypothesis is that the model histograms resemble the observed histograms and that this resemblance is meaningful. By “meaningful” we understand that the resemblance is the specific result of the parameter values chosen. According to the null hypothesis, the resemblance is accidental, which, in this particular case, implies that any arbitrarily chosen set of parameter values could have produced a comparable resemblance.

The null hypothesis was specified as follows. Remember that the model uses three parameter sets, i.e. sets of parameter values that are considered specific for the three types of dyads. The random model employs randomly chosen sets. However, since a randomly chosen set can in principle contain any possible and eventually highly unrealistic value, the selection was confined to specific parameter limits. We arbitrarily set these limits to *twice the range of the parameter values* featuring in the real (i.e. specific) model\(^9\). Note that in this case the model parameters for the

---

\(^9\) For instance, in the first parameter group (strength of concerns) in the model, the rejected child in the "rejected" dyad has the highest value, namely 0.8, for the concern Involvement (CI), the average child in the “average” dyad has a value of 0.7 and the popular child in the “popular” dyad has the lowest value, namely 0.6. The range between the highest and lowest value equals 0.8 – 0.6 = 0.2. The random model is based on an arbitrary choice of parameter values within twice that range, i.e. between 0.5 and 0.9. Comparable ranges for random parameter selection are of course determined for all other parameters.
three types of dyads are nested within this broader range.

The statistical method is as follows: First, a measure of fit (chi) is calculated, which is the sum of the distances (absolute difference) between the proportion of cases in each model bin and the corresponding empirical bin. The same is done for the random model. The difference between these two chi’s is an indicator for the relative fit quality of the model. We expect a considerably better fit for the real than for the random model, i.e. we expect a great difference in favor of the real model.

Table 4.5. The results of fitting procedure over distributions; per variable, per status group, as expressed in p-values.

<table>
<thead>
<tr>
<th>type of dyads</th>
<th>rejected dyad</th>
<th>average dyad</th>
<th>popular dyad</th>
<th>randtotal (p&lt;0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>child variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directedness</td>
<td>0</td>
<td>0</td>
<td>0.28</td>
<td>2</td>
</tr>
<tr>
<td>Proportion shared directedness</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Positive expressions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Intensity positive time</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Intensity positive number</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Proportion shared positive</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Negative expressions</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Intensity negative time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Intensity negative number</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>play partner variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directedness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Proportion shared directedness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Positive expressions</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
<td>2</td>
</tr>
<tr>
<td>Intensity positive time</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Intensity positive number</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Proportion shared positive</td>
<td>0.74</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Negative expressions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Intensity negative time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Intensity negative number</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>dyad variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coherence dyad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Shared positive expressions</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Shared negative expressions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Contrast dyad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>randtotal (p&lt;0.01)</td>
<td>20</td>
<td>22</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>per status group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. a. The fit quality is expressed in p-values: e.g. a p-value of 0 (p <= .001) for the variable ‘directedness child’ for the ‘rejected’ group of dyads means that the real model resembles the empirical distributions better than the random model, for this variable, for this group of dyads.
b. The rand total per variable expresses the number of times that p < 0.01 for this variable; the maximum value is 3 and the minimum value is 0. The rand total per status group expresses the number of times that p < 0.01 for this status group; the maximum value is 22 and the minimum value is 0.
Validating the agent model

The null hypothesis assumes that the values of the random model and the values of the model (each 5000 x) come from the same underlying distribution. This assumption is tested with the help of a random permutation procedure (1000 runs).

*Results based on p-values:* Table 4.5 shows the p-values resulting from the random permutation procedure.

To simplify the interpretation of this table, the number of times that the p-value is smaller than 0.01 has been counted for each variable (represented in the rows) and for each type of dyad (represented in the columns).

Overall, in 10 of the 22 variables, the real model resembles the empirical distributions better than the random model for all status groups (p < 0.01); in the remaining 12 variables this is the case in two of the three status groups.

For the rejected group of dyads, in 20 variables, the real model resembles the empirical distributions better than the random model (p <= .001). For the average group, the model is always better than the random model. However, for the popular group, in 8 variables the model does not give a better representation of the empirical data than the random model. In fact, in 7 out of 8 variables the random model gives a better representation of the empirical data than the real model.

Based on these findings with the p < 0.01 criterion, we consider the fit ‘excellent’ for the “average” dyads (22 variables out of 22 are significantly better); for the “rejected” group the fit is considered ‘very good’ (20 variables significant). Finally, we consider the fit for the popular group ‘moderate’, because of the combination of a low number of significant p-values (14) with a considerable number (7) of high p-values that refer to a better fit of the random model.

For a discussion of the fits resulting from a random model that is not limited to realistic values, we refer to the Endnote.

*Sensitivity analysis*

The question is to what extent the model reacts in case of changes in values of input parameter groups. This question relates to the point of limits of parameter ranges answered in the previous section, and concerns the extent to which the model is too sensitive or too insensitive for changes in parameter values. Figure 4.2 shows examples of three degrees of fit sensitivity, based on imaginary cases, and one empirical example.

An optimal fit occurs, if a relatively small fraction of the parameter interval corresponds with the empirical confidence interval (figure 4.2a).
Figure 4.2. Examples of three degrees of fit sensitivity, based on imaginary cases, and one empirical example: a. Sensitive parameter model; b. Oversensitive parameter model; c. Insensitive parameter model; d. Empirical parameter model.
**Exploratory qualitative analysis of sensitivity**

The question is whether the model is either too sensitive to small or too insensitive to greater changes in parameter values.

**Criterion: oversensitivity**

A model is too sensitive if a small change in a parameter value causes a major change in the fit of the model. Figure 4.2b shows that small parameter changes produce major changes in the fit quality. We have no reason to assume that this takes place in reality. For instance, it is highly unlikely that a very small increase in the Involvement concern parameter value would lead to a major difference in one or more of the variables, leading to a very rapid deterioration of the fit, whereas a big increase would again suddenly lead to an optimal fit. Especially in stochastic models (in which much randomization is used), the standard expectation is that no over-sensitivity will occur. Conversely, the expectation is that the model will behave correctly, showing s-shaped or comparable curves, in which monotonous increase can be seen within a sufficiently large range.

The question is operationalized as follows: does the model show too many fluctuations in its output variables, if one parameter at the time is varied from a minimum value towards a maximum value? Note that because it is virtually impossible to calculate all possible combinations of parameter values, we confine ourselves to manipulating one parameter at a time, and study the corresponding figure.

**The statistical method:** Charts were generated with on the x-axis a parameter that varied from a minimum value to a maximum value; and on the y-axis the accompanying model output for a particular variable. Charts were visually inspected with a focus on discontinuities and fluctuations that qualitatively resemble figure 4.2b.

**Results:** We visually inspected a random selection of parameters, namely ‘Involvement of the child’, ‘Involvement of the play partner’ (both belong to the Concerns parameter group), and the ‘contribution of the behavior to the realization of the concerns’ (Realizability parameter group), and checked the associated behavior of six important variables. The six variables were the dyad’s ‘coherence’, the child’s ‘directedness’, ‘positive expressions’, ‘intensity positive time’ and ‘negative expressions’, and the play partner’s ‘directedness’. There occurred no discontinuities, instabilities, and fluctuations in the figures that could point to oversensitivity of the model. Figure 4.2d presents an illustration of this finding, namely in the variable ‘coherence dyad’ with the parameter ‘involvement of the child’, for the rejected group of dyad. It
shows that an s-shaped curve appears, with a small dip on the right side.

**Criterion: insensitivity**

The question whether the model is insensitive can be answered by checking whether a major change in a parameter value causes no noticeable changes in the fit. Figure 4.2c gives an example of an imaginarly insensitivite model, namely in almost the whole range of the parameter value (from minimum to maximum value), the variable has an optimal fit.

The question is operationalized as follows: does the model show too little fluctuations in the course of the variables, if one parameter at the time is varied from the minimum to the maximum value?

The statistical method is the same as described above. The only difference is that we now check whether a variable remains within the limits of the empirical confidence interval when the underlying parameter is varied.

**Results:** We visually inspected the same sample as described in the previous paragraph. We noticed that in none of the variables the range of parameter values that yields a good fit is too broad; i.e. the model is not insensitive.

**Using sensitivity analysis to determine the quality of fit**

The question is whether sensitivity analysis can be used to determine the quality of the fit.

**Statistical method:** First, we checked whether the simulated average of a certain variable falls within the limits of the empirical confidence interval. We did this for discrete subranges (bins) of the parameter in question. Each time a positive result occurred in the bin, the similarity yielded one point. We did this for 6 important variables in total. All 1’s for each bin are summed, and the total number represents the fit of the model for this parameter, in this particular bin (see figure 4.3).

**Results:** As an illustration, figure 4.3 shows the parameter ‘child_involvement’, and six variables (three variables about action, 3 variables about emotional expression), of the “popular” group of dyads.

We can see that if the parameter ‘child_involvement’ has a value between 0 and 0.43, only 1 variable (‘directedness play partner’) falls within the limits of the empirical confidence interval. Between 0.43 and 0.5, none of the variables gives a good fit. However, between 0.5 and 0.57, five variables fall within the empirical confidence interval, which means that the fit is good, but not optimal (fit quality 5). After this, the fit quality decreases, to 3, 2, and finally 1.
Validating the agent model

Notice that the good fit we found between parameter values 0.5 and 0.57 corresponds with the parameter value in the model for the popular child in the “popular” dyad, which is 0.57. This latter value has come about in a fairly intuitive process of estimation of values, by means of inspecting graphs, simulations, and fitting averages over the total group of dyads. This means that the sensitivity analysis eventually provides a good underpinning of the estimation of parameter values.

DISCUSSION

The goal of this chapter was to report about the empirical validation of a general explanatory model of short-term social behavior in a developmental context. This model was based on dynamic systems principles, and was applied to interaction of children of various sociometric statuses during play. Overall, the model showed a good empirical fit.

Implications for explaining social interaction of six-and seven year-old children of different sociometric statuses

We can consider the good fit of the model as support for the model’s theoretical assumptions, such as the notion of goal-directedness of interactions, in which

Figure 4.3. Sensitivity analyses of the parameter ‘child_involvement’, and six variables (three action based variables, and three expression based variables) in the ‘popular’ group of dyads.
concerns are the guiding force behind a child’s actions and emotional expressions. In addition, the model underlines the role of emotional appraisals, pleasure and non-intentional aspects of behavior, such as contagiousness in an interaction process. Moreover, it supports the idea that social power and social effectiveness are differential aspects that play a crucial role in explaining differences in social interaction of children and dyads of various sociometric statuses. The model describes interaction as an iterative process, in which both the child’s own and the play partner’s properties and behaviors dynamically intertwine. That is, at any moment in time, one child affects the other and the other child affects the first, thus creating the conditions under which both participants will operate during the next moment in time.

Second, the model helps to better understand empirical findings in specific contexts, and thus, potentially contributes to better prediction and intervention. A first empirical finding is that specific differences and similarities do exist between popular, average and rejected children and dyads.

An example of such a difference is the finding that - in this specific interaction setting - rejected children show more ‘positive expressions’ than popular children. Our model helps to understand this finding by referring to the major role of concerns. It describes the specific process in which these children use their emotional expressions in trying to establish satisfactory interactions with others, and the specific way in which other children react to these expressions. This leads to rejected children showing an ‘overflow’ of positive expressions.

This finding is related to the finding that the amount of positive expressions that are reciprocated by a positive expression of the play partner (as expressed in the variable ‘proportion shared positive’) is higher for popular children than for rejected children. Our explicit computational model explains this by showing how the child’s social effectiveness and social contagiousness act as mutually amplifying forces in the case of high power individuals. These predictions could probably not have been made on the basis of existing models, which are considerably less context-sensitive and not process-based.

A better understanding of empirical phenomena should lead to better intervention and assessment, i.e. should have practical and applied consequences. That is, first, in addition to focusing on child specific factors, assessment and intervention should also explicitly account for context-specific factors in a process-oriented framework. Desired changes in behavior can be realized by varying the context. For instance, the teacher can let a rejected child cooperate with a more popular child in the
Validating the agent model

class. This must be done in a particular action context, in which the rejected child can practice new skills, and experience positive emotions in doing so. In addition, it is important to see interventions as part of an ongoing iterative process, in which each step in the process helps to determine the next intervention step. Furthermore, with regard to assessment, knowledge about concerns and appraisals of a child can be an important tool for understanding the child’s behavior and for achieving a successful intervention. For instance, before involving a child in a social skills training, the child can be interviewed about his concerns with regard to his socially undesired behaviors, such as aggression (see Singer, Doornenbal, & Okma, 2004; in which the same kind of interviews are described for foster children).

Third, the model is a convenient tool for (computationally) experimenting with various assumptions - for instance, about the height of the concern involvement in rejected children - by changing parameter values, and observing the associated changes in outputs. In addition, the model shows how group characteristics follow from individual, in this case, dyadic processes. For instance, insight can be obtained in the potential causes of intra-individual variability in patterns of behavior of one individual child. Other possibilities concern the simulation of conditions in which more or less variability between individuals and / or groups occur. This allows the researcher to obtain a better insight in the bandwidth of possible effects of varying parameter values.

A point of consideration concerns the moderate fit of the popular group with regard to the distributions per variable separately. This lack of fit could be explained by the way in which we have incorporated the social effectiveness component of children in the model (see the more extensive description on the website www.gmw.rug.nl/~model/, which explains that social effectiveness is not implemented in parameter groups that relate to emotional expressions). In additional research, (see chapter 2; Steenbeek & van Geert, 2005b) we found that the child’s social effectiveness plays a major role in realizing concerns. A high level of social effectiveness is particularly characteristic of popular children.

Methodological implications regarding the fitting procedures

A first implication concerns Murray’s relativizing the importance of empirical fits. He states that “fitting the data does not tell you that you have the right mechanism…….We have to free ourselves from the idea that goodness of fit is the sine qua
non of science” (Gottman et al., 2002; p. 67-68). This statement supports the necessity of a good theoretical justification of a model. The empirical fit makes sense only against the background of the underlying theory. We have tried to conform to this requirement by first, incorporating the process component in the model, in particular by using principles from dynamic systems, and second, by founding the model on adequate theoretical assumptions.

A related issue is that the number of parameters in the model follows directly from the underlying theory, which in the case of our model has led to a considerable number of parameters. This number is not primarily determined by a statistical criterion, in which the model is penalized for the number of parameters used. (Examples of such goodness of fit criteria are Akaika’s Information Criterion (AIC) and Bayesian Information Criterion (BIC)).

On the other hand, it does not seem very plausible that all dyads from the same dyad group have the same parameter values, as is now implicitly assumed by the model. It is more likely that each type of dyad is represented by parameters that consist of a range of values, instead of just one specific value. Testing the fit qualities of such ranges could be a next step in the further elaboration of the model.

A last remark concerns the use of a random component in the model. This random component is incorporated in the model in such a way that its influence is more or less constant, without dominating the influence of the values of the input parameter groups. However, this model random component can be no more than only an approximate representation of the empirical random component, in which all kinds of variables exert an influence of varying strength. In further research the magnitude of the influence of this random component could be estimated, comparable with the estimation of error variance in more standard models.

**Implications regarding modeling in developmental psychology**

An implication of using such a complex model instead of a simpler one is that we can make use of the additional value of the complex output of the model, in the form of interaction patterns unfolding in real-time. That is, the model yields much ‘richer’ information than merely producing an equilibrium state or an average, and thus leads to detailed, testable hypothesis. The question why we did not make use of this ‘richer’ source of information in validating the model can be answered by emphasizing that performing fitting procedures on a group level is a first, conditional step.
Validating the agent model

before fitting on the level of individual patterns can be performed. Now that this first step yields encouraging results, the next step - validating the model with regard to patterns that emerge within one child, within one interaction – can be taken (as described in van Geert & Steenbeek, 2005).

A second important implication of this short-term model concerns its function in the explanation of long-term development. Basically, the short-term dynamics of a play session or other form of interaction that involves dyadic interaction can be conceived of as a single step in a time series of such interactions over a long term, for instance a school year, or a developmental period such as childhood. The long-term development concerns the long-term change of the parameter values that govern the short-term model. For instance, it is possible that in a child who encounters many unsuccessful interactions with others and who thus cannot sufficiently realize his concern for pleasurable interactions with others, a slow decrease of the involvement concern parameter will occur, leading to diminishing interest in others and thus to a more or less self-caused rejected status.

In order to expand the short-term model and make it into a model over the long-term, it would be necessary to add a new component to the model, namely a social network of potential interaction partners. This long-term model must of course specify how a preceding interaction affects a successive interaction. For instance, how is the Involvement concern-parameter that is associated with a potential interaction partner, updated after an interaction with that child in question, dependent on how pleasurable the past interaction was.

It goes without saying that although the use of a short-term model in order to understand long-term development is not strictly necessary, it nevertheless provides a deeper and more complete explanation of the developmental process at issue. In addition, if one wants to intervene in a problematic social development – e.g. consisting of many unsuccessful interactions with others – it is very helpful to have a model of the real aggregation level at which changes will take place (for more information about the coupling of a short-term model and a long-term model, we refer to van Geert & Steenbeek, 2005).

Third, a more common approach to explaining social interaction in a developmental context is to measure a set of independent variables (e.g. concerns) in a particular sample of children, to measure a set of dependent variables (e.g. amount of positive expressions), and then calculate an association between the two (e.g. common variance). We have reasoned that such independent variables are in fact the
product of an ongoing process (dyadic interaction) and thus, that a model predicting those variables should be a model of the process that produces them. However, this process model introduces a number of assumptions that are incompatible with the notion of independent variables. It remains an important challenge for further research to explore the possibilities of a (relatively) independent measurement of the main parameters of the model. For instance, it is likely that the child’s concerns for ‘playing together’ with particular peers are consistent with the child’s sociometric ratings, provided they are taken shortly before videotaping a play session.

A last consideration refers to using simulation models in order to generate predictions, and validating such models with empirical data, in the field of developmental psychology research. Without doubt, this research method is still in its infancy, notwithstanding the fact that important work has been done already by researchers such as Thelen & Smith (1994; about motor development), and Gottman (1997; 2002; about interaction processes). However, especially in the field of research of social development, there are relatively few examples of studies that combine modeling with the development of tools and steps for fitting procedures, and for determining the quality of the resulting fit. This lack of examples contributes to making our research a highly exploratory process.

As a concluding remark, we think that the scope of the model’s basic assumptions and possibilities is such that it can also be used to study other forms of interaction processes, such as those taking place in groups of children, and between children and adults. Hopefully, the model will serve as an initiative for future use of simulation models as tools for studying the ‘grammar’ of behavior, and by doing so help us to better understand the mechanisms of short-term behavioral processes and thus contribute to our understanding of long-term social development.
Validating the agent model

ENDNOTE

Fitting distributions per variable

Results based on an alternative random model (with a non-overlapping parameter range)

The question is: if qualitative (dis)similarity between model and random model is considered important, does it matter whether one uses specific parameter values, for instance based on theoretical assumptions, or will randomly chosen parameter values also suffice? Remember that in the random model that just has been tested, the range from which parameter values were randomly sampled, shows a 50% overlap with the range covering the parameter sets based on a deliberate, theoretical choice (see the aforementioned explanation of taking twice the width of the theoretically justified range). We have just seen that this particular random range results in distributions that are qualitatively similar to those of the real parameter range, at least for a significant number of variables. If the values of the parameters do not matter, a comparable qualitative similarity should also result if we take parameters from an entirely different parameter range, for instance, one which is also twice as broad as the theoretically determined range, but which is moved to the left (i.e. covering the lower part of the possible parameter values). Note that the model of the three types of dyads is not nested in this alternative model.

Statistical method: is the same as described in the paragraph ‘testing an arbitrary model’.

Results: All variables yield significant p-values (p < .001), except for one variable (‘negative expressions’ for the group of “popular” dyads). This means that the real model distribution resembles the empirical distribution better than the random model for all variables but one and for the three types of dyads. This picture is confirmed by visual inspection of the data, which is illustrated in Steenbeek & van Geert (2005a). The difference between the real model and the alternative random model is much bigger than the difference resulting from the original random model. Thus, the question whether any randomly selected set of parameters can achieve a reasonable qualitative fit with the empirical distributions can be answered with “yes” if the parameters are selected from a range of values that overlaps with the theoretically funded range, and with “no” if the parameters come from a more peripheral, non-overlapping range.