Explaining Sociometric Status Classification Results via Cognitive Modeling

Terrence C. Stewart (tcstewar@connect.carleton.ca)
Institute of Cognitive Science, Carleton University
1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

Robert L. West (robert_west@carleton.ca)
Institute of Cognitive Science; Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

Robert Coplan (robert_coplan@carleton.ca)
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

Abstract

A series of computational models are presented which address the question of how peer relations change over time. We examine data from a standardized metric (CDC) which places school children in one of five categories: Popular, Rejected, Neglected, Controversial, and Average, and how such classifications change over time. Our models are created to match this data, and are then used to predict the how this classification changes due to different sorts of traits that can vary per individual. Of these predictions, some have been confirmed by turning to standard psychological results, while others require further investigation. Both of the successful models exhibit highly similar behavioural patterns, even though they are fundamentally different sorts of models. We argue that this indicates both models are capturing an underlying regularity of the social dynamics.

1. Introduction

Popularity is generally thought of as something dependent on who we are: on our personality and social skills. However, in this paper we examine the opposite proposal: that popularity has very little to do individual behavior. To do this, we take a dynamic, multi agent approach to modeling popularity. That is, we examine how popularity is distributed across time within a dynamic network of interacting agents. As Sun (2001) notes, multi-agent models exist across a variety of disciplines, but there is very little discussion of what constitutes a valid multi agent model (although see Sun, 2006 for a start on this problem).

There is a tendency to think of popularity as something that resides within a person, but it is actually a distributed property of the group. An individual's popularity is a function of how many people like them. The notion that popularity resides in a person may be a result of viewing popularity as the outcome of our personality and/or social skills, so that when someone says that a person is popular, what is meant is that the person possesses the means to be liked. In contrast to this view, we propose that popularity is an emergent property of the dynamics of the group interaction. This means that where an individual ends up in the social environment depends on where they are deposited by the forces generated by the dynamics of the group interaction.

To investigate this we used multi-agent models of children's friendship interactions and compared the results to the empirical data from studies in this area. In particular, we were interested in the relationship between the emergent properties of the system as a whole and the cognitive properties of the individual. Since there is no normative method for doing this (see Sun, 2006 for examples and discussion), we used two principles of cognitive modeling to guide us. First, the *simplicity principle* has been put forward as a fundamental principle of cognition by Chater (2003). Essentially, the idea is that the brain tries to do things in the simplest way possible. As Chater points out, a wide range of cognitive phenomena seem to conform to this principle. We used this as a basis for creating our models, leading to the creation of the simplest possible cognitive model for processing friendship information. Second, the *architecture principle* is based on Newell's concept of cognitive architectures (Newell, 1990). Essentially, the idea is that all the different things that we do have to be performed by the same brain, and therefore the same cognitive architecture. Thus, while it is useful to create unrelated models of different cognitive abilities, ultimately it should be possible to implement all of those models in a single architectural system that, following the simplicity principle, is as simple as possible. This guided us to further develop models using ACT-R (Anderson and Lebiere, 1998), which is a cognitive architectural system that has been used to accurately account for hundreds of different cognitive effects.

We first developed a simple computational model of the process of friendship formation, and then, after examining this model's behavior as compared to real-world results, re-implementing it in ACT-R. We require that it be straightforward to do this implementation in ACT-R (as per the simplicity principle) and that the ACT-R version also account for the human data. Interestingly, the final ACT-R model suggested a qualitatively different explanation than the simple model.

1.1. Background

Developmental psychology has made it increasingly clear that external factors are key to understanding how we become who we are. Indeed, even for something as important as being *popular* (i.e., "well-liked" by peers), internal factors seem to only explain approximately 30% of the observed variance (for example, see Rubin & Mills, 1988). More generally, even factors as seemingly important as self-esteem have been shown to have surprisingly little correlation to lifetime success (Baumeister et al, 2003).

Individual factors do influence popularity; a child could be unpopular because they are shy, aggressive, and/or socially incompetent (Newcomb, Bukowski, & Pattee, 1993). However, the proportion of the variance explained by these internal factors tends to be small. Indeed, in (Rubin and Mills, 1988), the R^2 values tend to be in the 0.25-0.30 range, with one exception at 0.43. This indicates that only 25-30% of the causes of differences in social status can be directly attributed to such individual features. The rest may be due to a variety of external factors in the dynamics of the group interaction.

The major difficulty in addressing this problem is the complexity of the individual's interaction with other individuals and the resulting potential for complex feedback loops. This complexity makes it difficult to predict how a particular event may impact an individual's life, and how an individual's choices will eventually impact them.

One methodology for addressing this type of situation is to develop multi-agent models. The goal in developing these models is the same as that of developing any scientific theory: to explain and predict behavior. The model is created based on a certain set of known empirical results, and then, once a suitable model is found, it can be used to produce predictions of other aspects of behavior. These predictions can then be empirically tested. Once a sufficiently accurate model is created, it is possible to perform exploratory research on the model. For example, one could investigate the impact of a particular intervention technique on the model before attempting it on real people.

In this paper, we present our work developing a model of peer relations over time within school-aged children. Our models were constructed to predict the same popularity categories of used by standard sociometric measurement techniques. Importantly, the models allowed us to observe how these relations change over time, and allowed us to investigate what impact internal factors can have.

1.2. Existing Research

The study of the measurement of children's peer relations (i.e. *sociometric classification*) begins with Moreno's (1934) research. His work on describing individuals in terms of the *attraction* and *repulsion* felt by them towards others and by others towards them spawned a wide variety of measurement techniques, each attempting to develop a useful scale for the investigation of the causes and effects of children's social experiences (see Cillessen and Bukowski, 2000 for a review).

One of the results of this research is that a simple, one-dimensional scale is not sufficient to capture useful information. In particular, we generally need to distinguish at least three categories: individuals who are viewed positively by peers, individuals who are viewed negatively, and individuals who are ignored. To do this, most modern sociometric techniques involve distinguishing two measurement dimensions: *preference* and *impact*. A person who is ignored would be one with a low impact, while someone who is actively disliked might have a high impact, but very low preference rating.

The most popular and widely used measurement scheme is known as CDC Classification. It is named after its creators, Coie, Dodge, and Coppotelli (1982), and classifies people into one of five categories: *Popular*, *Rejected*, *Controversial*, *Neglected*, or *Average*. To facilitate its use across as wide a range of ages and situations as possible, its methodology is quite simple. Using interviews or questionnaires, each person is asked to name three people in their peer group that they like, and three people that they dislike. These scores are than standardized within class to control for the number of possible nominations received. The simplicity of this measurement is important for measuring popularity in young age groups. Using the survey results, each individual is given an *Acceptance* score (the total number of times that person is listed by other people as someone they like) and a *Rejection* score (the number of times they appear on the 'dislike' lists). A *Preference* value (Acceptance minus Rejection) and an *Impact* value (Acceptance plus Rejection) are also created, where Preference refers to whether you are more liked or disliked and Impact refers to how much people pay attention to you. Individuals are then classified into the five categories according to the rules shown in Table 1.

Table 1: The decision rules for classifying with CDC. All values are normalized to a mean of 0 and a standard deviation of 1.

Category	Rule
Popular	Preference>1 Acceptance>0 Rejection<0
Rejected	Preference<-1 Acceptance<0 Rejection>0
Neglected	Impact<-1 Never appears on anyone's 'like' list
Controversial	Impact>1 Acceptance>0 Rejection>0
Average	None of the Above

Given the wide use of this system, and the accompanying availability of a wide variety of experimental results that make use of CDC classification, we decided to use it as the basis of comparison for our computational modeling results. It is worth noting that other methods do exist in the literature (such as Newcomb and Bukowski, 1983), and could be used for future investigations into the models of peer group relations that we have developed.

1.3.Comparison Data

Before describing the models we developed, it is useful to examine the real-world data that we used for comparison. Newcomb, Bukowski, & Pattee (1993) provide complete results from nine different studies that used CDC categorization on a total of 2,571 students, ranging from kindergarten to grade 9, in groups of around 30 individuals. Overall, these nine studies give the following 95% confidence intervals for the percentage of people in each category (see Table 2). This data set gave us our first basis of comparison between the model and reality. Any successful model must give results within these ranges.

Table 2: The 95% confidence intervals indicating the percentage of people in each CDC category.

Popular	Rejected	Neglected	Controversial	Average
7%-32%	12%-26%	0-28%	1.6-16%	5.9-69%

As a more stringent test of our models' abilities, we turned to the measure of categorization stability over time. Cillessen, Bukowski, & Haselager (2000) give the results of a meta-study which collected the change in CDC categorization over periods of time ranging from one month to four years. This tells us what percentage of individuals who were put in a particular category at one point in time, remained in that category in a later point in time. Table 3 gives the 95% Confidence Interval for this data.

Table 3: The 95% confidence intervals for the stability of each CDC category.

Popular	Rejected	Neglected	Controversial	Average
33-44%	39-49%	20%-30%	24%-36%	51%-69%

A striking result from this study was that the stability values did not seem to be affected by the length of time between measurements. That is, if someone is currently classified as Rejected, they have a 39-49% chance of being classified as Rejected one month from now *and* four years from now. This independence between stability and time has important implications for the models being developed. Since the models are of individual behavior over time, the longer the model is run for, the more simulated 'time' is passing. If we keep performing stability measurements, then the stability should approach the values shown in Table 3. Importantly, this meant that we did not need to figure out how time within the model corresponded to real-world time. Instead, we required that the model not only match the stability data, but also give this result if run for a sufficiently long length of time.

For all of our models we used a virtual class size of 30 agents, to roughly approximate the average class size used in the research on real children.

2. The Random Model

To explore the CDC classification system, we began with two extremely simple cases. In the first case, we assumed that the same three people are always chosen as liked, and that the same three people are always chosen as disliked. This represented the extreme case of the popularity categories being determined solely by stable individual differences. That is, everyone likes the same three people the most and everyone dislikes the same three people the most. As Table 4 shows, this produces a pattern of results *inconsistent* with the known real-world category sizes, and thus is clearly not a good model of popularity.

Table 4: The category distribution for a model where the same people are chosen as 'liked' and 'disliked' by everyone in the group, for a group size of 30.

Popular	Rejected	Neglected	Controversial	Average
10%	10%	0%	0%	80%

In the second case, we assumed that everyone has an equal chance of being chosen as liked or disliked. This represented the extreme case of the popularity categories being determined randomly. That is, nobody has any stable preferences. To test this we ran a simulation where the agents simply chose their answers at random (with the condition that they could not choose the same person twice). Interestingly, these results (shown Table 5) did not fall outside the confidence intervals for the categorization data.[1]

---

[1] This phenomenon of purely random data giving similar sizes for each CDC category has been noted previously (Cillessen & Ten Brink, 1991; Maassen et al, 2005). However, instead of using this as an argument for abandoning CDC classification, we prefer to simply note that since a simple (random) model can produce this data, then a more complete model should be able to produce this result *along with other results as well*.

Table 5: The category distribution for a model where people are ‘liked’ and ‘disliked’ at random. The group size is 30, and results are similar for other sizes.

Popular	Rejected	Neglected	Controversial	Average
12%	12%	7%	2%	67%

Taken alone, this result suggests that the friendship formation process could be roughly approximated by a simple random model. However, if we also consider the data on category stability (shown in Table 3), then this model can also be shown to be false. Because the random model assumes no continuity across time, the chance of remaining in a group is the same as the chance for being put in the group, so the stability percentages match the category membership percentages (i.e. the stability of the Controversial group is 2%, and the stability of the Popular group is 12%, and so on). This is well outside the confidence interval for the human data, and so the random model can also be rejected.

However, the fact that the random model can approximate the category membership results allows for an interesting characterization of the friendship process. It implies there is no reason to assume that individual differences make a strong difference for getting into a category, but once in that category there may be a mechanism that retains members at a greater than chance rate. This is consistent with the finding that individual differences do not account for a high percentage of category membership.

3. The Computational Model

As indicated previously, the simplicity principle suggests the creation of models that are as simple as possible while still accounting for the phenomenon of interest. After examining the capabilities of the random model (which was successful at modeling some, but not all, of the relevant phenomena), we tried to develop the simplest model that would have the following human characteristics:

1. Each person should remember how much they like each other person
2. People should use this memory to determine how they will ‘interact’ with others
3. People should use the results of this ‘interaction’ to change how much they like the person they just interacted with

This led us to the following algorithm:

```
1.Let a[i,j] be the amount person i likes person j
2.For all pairs of individuals i,j:
    a.Use a[i,j] to determine how i behaves towards j (referred to as b[i,j])
    b.Use a[j,i] to determine how j behaves towards i (referred to as b[j,i])
    c.Update a[i,j] based on how j behaves towards i
    d.Update a[j,i] based on how i behaves towards j
3.Repeat step 2 for as much time as is desired
```

It is important to note that we are modeling the ‘positiveness’ or ‘negativeness’ of an individual’s behaviour with a single scalar value. This means that a highly positive action is represented by a number like +10, and a slightly negative action would be something like –0.7. We do the same thing for how much an individual likes someone else. We can thus imagine the following scenario in the model:

> Two individuals (X and Y) are interacting. X likes Y a lot (+10), and Y likes X a little bit (+4). X chooses to act very nicely towards Y (+9), and Y chooses to act somewhat nicely towards X (+5). After the interaction, X and Y will change how much they like each other to new values (+8 and +6).

To implement this as a specific model, we needed to define an algorithm for each agent to use to decide how to behave towards each other, based on how much they like each other. We choose a simple method: generate a normally distributed random value with a mean of a[i,j] and a standard deviation of 1. This value represents how 'nicely' individual i is going to behave towards j.

$$b[i,j] = a[i,j] + N(0,1) \qquad \text{(formula 1)}$$

It turns out that changing this deviation does not affect the overall behavior of this model, so we leave it set to 1 for the moment. We will examine what happens if we allow the individuals in the simulation to have different deviations later in this paper.

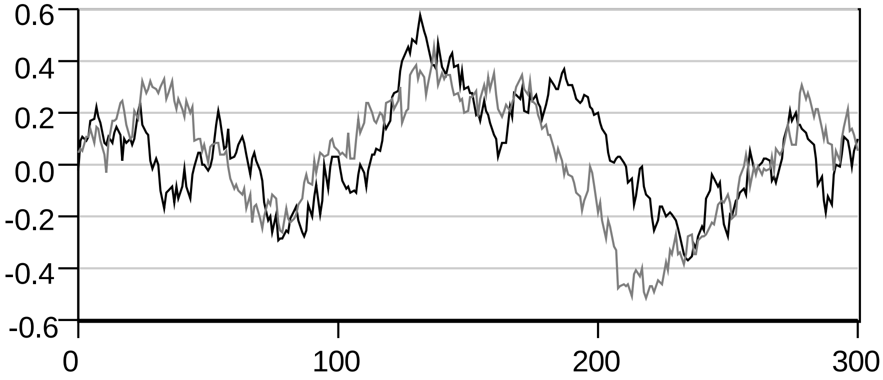
Similarly, we need to update how much i likes j, based on j's actions. One idea would be to simply add b[j,i] (how 'nicely' j has behaved toward i) to a[i,j]. However, this approach can be shown to cause a positive feedback loop which would mean that if i currently likes j and j currently likes i, they will keep increasing how much they like each other to increasingly large values. This is clearly does not capture the ebb and flow of real human relations. Instead, we use a slightly more complex formula which acknowledges the role that a[i,j] (the amount person i likes person j) could play in determining how person i would react to the behavior of person j.

$$a[i,j] \leftarrow a[i,j] + r(b[j,i] - a[i,j]) \qquad \text{(formula 2)}$$

The idea behind formula 2 is that individuals in the model evaluate the actions of others according to the expectations they have for them. This is based on the insight that we expect a friend to be friendly and an enemy to be unfriendly, and we are not surprised when this happens (and thus would not change our beliefs). However, if a friend were to be unfriendly we could be quite hurt, and likewise, if an enemy were to be kind it could surprise us and could cause us to reevaluate our feelings toward them. Interestingly, this formula is a simple estimator, which means that a[i,j] can be thought of as an estimate of b[j,i]. In other words, how much an individual in the model likes someone else is an estimate of how that other individual will behave toward them.

The long-term consequences of following even this simple model are not obvious. Since repeated iterations of this model will cause a[i,j] and a[j,i] to interact with one another, it is instructive to view what typically happens to these variables over time. Figure 1 illustrates the effect. The results indicate that while the two individuals in the simulation generally like each other by about the same amount, there is a considerable amount of change going on that could potentially capture (at least quantitatively) the ups and downs of real human relationships.

Figure 1: Two individuals interacting over 300 iterations of the model. Shown is a[i,j] and a[j,i]. r is 0.05.



It should be noted that in doing this we have introduced a parameter into this model: r. This is the ‘adaptation rate’ and controls how quickly an individual changes its opinions of others. It can be considered to be similar to the 'learning rate' parameters found in a wide variety of other computational models, such as neural networks (Rumelhart et al, 1986), ACT-R (Anderson & Lebiere, 1998) and the Rescorla-Wagner model of Pavlovian conditioning (Rescorla & Wagner, 1972).

3.1. Category Distribution Results

It turns out to be easy to show that this new model is exactly as good as the random model at predicting the distributions of categories. We first note that, given the preceding algorithm, the values a[i,j] and a[j,i] change in a manner that is independent of anything else in the simulation. This means that when the CDC nominations occur (i.e. when we check each individual to see which three others they like and dislike the most), the chance of a particular person j nominating another person i is still independent of person k also nominating person i. This in turn means that, for any given CDC evaluation, the number of nominations an individual receives (both positive and negative) will have the same sort of distribution as the initial completely random model. This predicts that we will continue to have exactly the same distributions of categories in this new model as under the random model. We ran the simulations for confirmation, and found that we did, in fact, get the same distributions. This result holds for all values of r and offers a reasonable account of why the random model works for the category membership results.

3.2. Category Stability Results

We can now determine how well this model predicts the category stability data. Our first challenge was to determine how to measure stability. Measuring stability consists of doing two CDC measurements separated in time. To do this in our model, we needed to decide on two different values: how much simulated time would occur between CDC measurements, and how much time would have elapsed in the simulation *before the first measurement*. This is because, when measuring the stability in the real world, the sociometric studies did not deal with children who had all just met for the first time. Instead, they were children who had been in a group situation long enough for friendships to form (at least months), who were then classified using CDC at two different points in time.

Neither of these values is straightforward to define. Fortunately, we do know from Cillessen, Bukowski, & Haselager (2000) that the stability values do not show a strong response to varying the length of time between measurements. That is, it does not matter much how much time passes between measurements; the result stays pretty much the same. This means that if our model is a good one, it should exhibit a similar behavior. To decide on how much time is spent before the first measurement, we took an empirical approach. We found that allowing the agents to interact for short periods of time initially (<100 time steps) lead to very low stability. This was somewhat expected, as the individuals were still ‘getting to know each other’ and settling into their roles. For the data presented here, we chose to use 1000 initial time steps to represent the amount of time the agent had “known” each other before popularity was measured.[2]

Figure 2 displays our model’s stability results for each of the five CDC categories. Each line in each graph represents a different setting for the adaptation rate (r). The x-axis indicates the number of iterations (i.e. time) between the two category measurements. The shaded area shows the 95% confidence interval from the real-life stability data.

---

[2] We also investigated much longer periods of time (10,000 time steps), and found that the only effect was to shift the range of suitable parameter settings for r slightly lower.

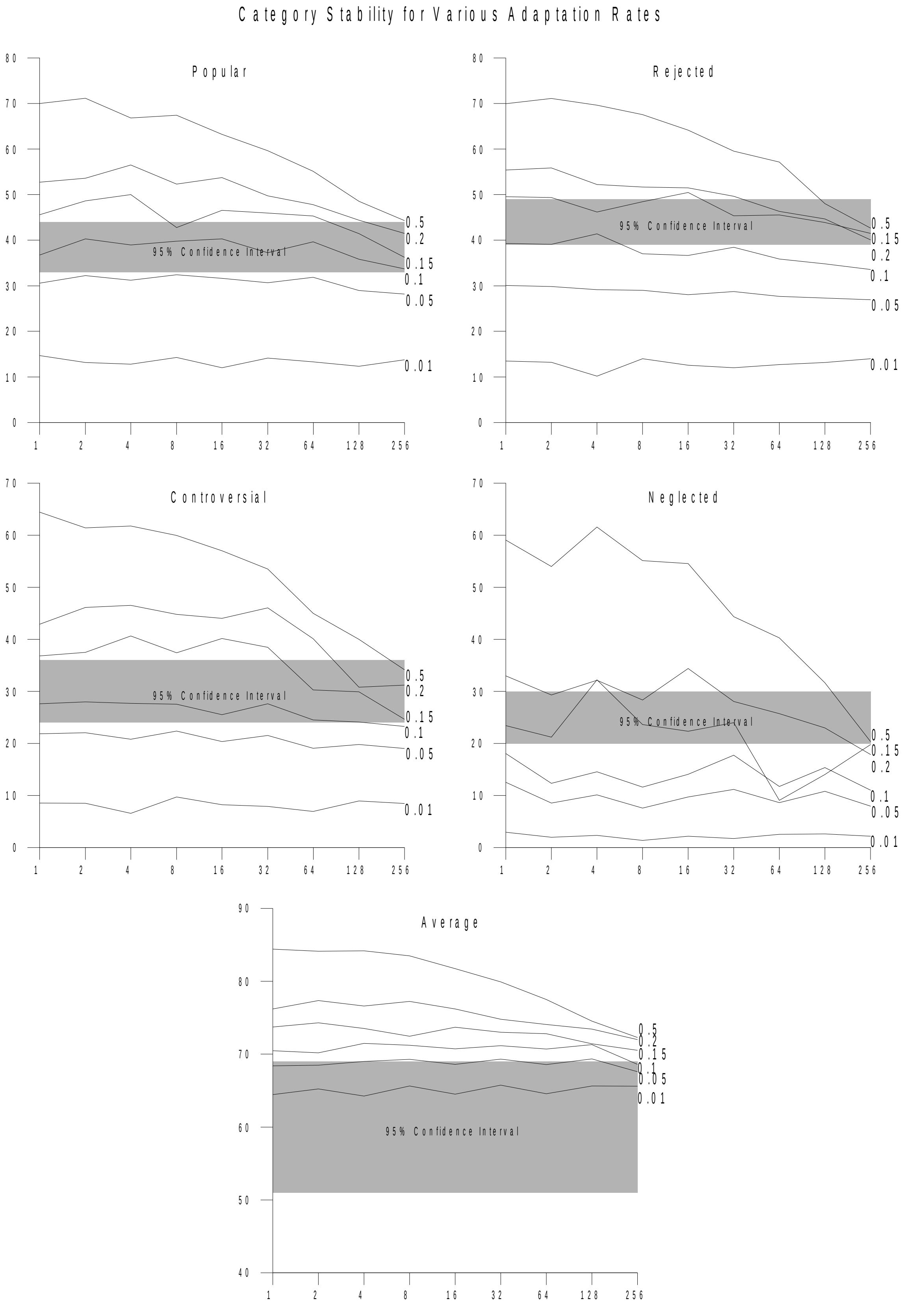


Figure 2: Stability results from the model (6 lines; one for each setting of the adaptation rate) compared to the 95% confidence interval for the real-world data.

The main result here is that the stability measurements match the expected results reasonably well for values of r around 0.1. We see that the stability tends to be within the confidence interval range, and this stability measurement is not strongly affected by how much time is spent between measurements (the slight downward trend is only noticeable due to the logarithmic time scale). However, it is true that higher values of r showed a slight downward trend in stability. It is hard to interpret this result, since the logarithmic nature of this effect means that it may not be noticeable in the time scales normally used in the human data (Cillessen, Bukowski, & Haselager 2000).

In any case, we have a reasonable behavioural match for a range of r values (0.05 to 0.2). Although 'parameter fitting' was used to find a good value for r, our criteria for a 'good fit' involved fitting five different measurements (the stabilities of the five categories), plus the constraint of not being strongly time-sensitive. These facts, plus the consideration that this is the *only* parameter in our model, allows us to be reasonably confident that it is capturing some important aspect of the phenomena.

3.3. Biasing Effects

The results shown thus far demonstrate that our model gives both category distribution and stability results that match the existing human data quite well. It does this over a wide range of settings for the one existing parameter within the model (r). Since this establishes our model as potentially useful for investigating the dynamics of popularity, we can now use it to do these investigations and produce some real-world predictions.

The domain we have chosen for this is the effect of individual differences. So far in developing our model, we have assumed that *every individual starts out identical*. This is clearly a false assumption. It is known within the sociometric literature that individual differences can account for a small, but significant portion of individual classification (for example, see Rubin and Mills, 1988). To investigate this factor, we systematically introduced different sorts of variations into our model to investigate four well-known effects. In all of these simulations, we allowed a particular aspect of each individual in the model to vary across a normal distribution. The size of this distribution was made as large as possible while still matching to the real-life data.

It should be noted that these investigations were done *after* we developed our model, and our model was not developed with these variations in mind. We are thus taking a model that was developed in one domain (the prediction of CDC stability data) and using it in a new domain (the effect of individual differences on CDC classification). This sort of model extension is the primary method that science uses for determining if a model or theory captures an important and useful aspect of the phenomenon in question. Any model that cannot be extended past the situation it was designed for may be merely descriptive, and thus not useful for understanding the fundamentals of the situation.

To begin with, we looked at the Hostile Attribution Bias. This refers to the fact that certain individuals tend to interpret the actions of others in a more negative light than is intended. Research has shown that rejected children (particularly those who are aggressive) are more likely to assume malevolent intent when they are faced with ambiguous social cues (Crick & Dodge, 1994; Dodge, Lansford, Burks, Bates, Pettit, Fontaine, & Price, 2003). To model this we inserted this individual difference into our model by adjusting formula (2) in the following way:

$$a[i,j] = a[i,j] + r(b[j,i] + B[i] - a[i,j]) \qquad \text{(formula 2b)}$$

In this new formula, $B[i]$ is an individual interpretation bias. This can be either positive or negative. Agents can thus be biased to be either overly negative ($B[i]<0$) or overly positive ($B[i]>0$) in how they interpreted the actions of others.

We then ran the simulation over 1000 groups of 30 agents each. Each agent had a value of B[i] chosen from a normal distribution with a deviation as large as possible while still matching the aforementioned results. After 500 simulation iterations, CDC classification was performed. We then measured the *effect size*. This is the number of standard deviations above or below the mean the values of B[i] are for the individuals within each category, as compared to the Average individuals. This is a technique used in (Newcomb, Bukowski, & Pattee, 1993) to relate individual variations in personality metrics to CDC classifications. We used this method was for all variations investigated.

The results in Figure 3 show that the simulation predicts that Rejected individuals tend to be overly negative in interpreting the behaviors of others, while the Popular ones tend to be overly positive. This is exactly in accord with the Hostile Attribution Bias. Interestingly, the model also generates a novel (and testable) prediction, as Neglected individuals also tended to be somewhat negative in their interpretations.

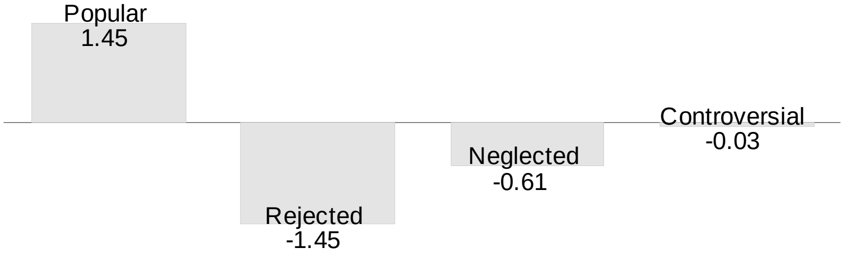


Figure 3: Effect size of varying interpretation bias (the average number of deviations a group is from the mean)

Another well-known finding is that Popular children tend to have good social skills whereas rejected children tend to have poor social skills. (e.g., Coie, Dodge, & Kupersmidt, 1990; Newcomb, Bukowski, & Pattee, 1993; Parkhurst & Hopmeyer, 1998). That is, popular children tend to express themselves in ways that mitigate bad feelings whereas rejected children tend to express themselves in ways that exacerbate bad feelings. To reflect this in the model we biased the mean of the Gaussian distribution used in Formula 1. Specifically, a positive value created a bias toward behaving nicely and a negative value created a bias toward behaving badly. Figure 4 displays the results and shows that the manipulation had the intended effect. Interestingly, the neglected agents were again shown to be somewhat similar to the rejected agents. Note also that varying the ability to express oneself socially produces approximately the same distribution of effects that varying the ability to accurately assess another persons' intentions (i.e., the hostile attribution bias). This makes sense as the end effect is essentially the same, that is to bias the valence of the behavior.

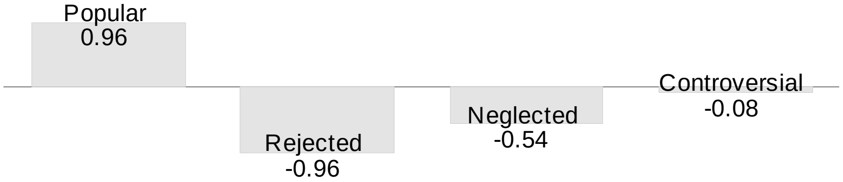


Figure 4: Effect size of varying behavior bias (the average number of deviations a group is from the mean)

The third effect we investigated was that neglected children have been shown to interact with their peers less frequently than average children (Dodge, Coie, & Brakke, 1982; Coie & Dodge, 1988). To reflect this in our model, we added an interaction probability for each agent. The percentage chance for two agents interacting was determined by multiplying their interaction probabilities together. As illustrated in Figure 5, this manipulation was successful in capturing the effect. It also produced the unexpected effect that a high level of interaction was associated with being Controversial.

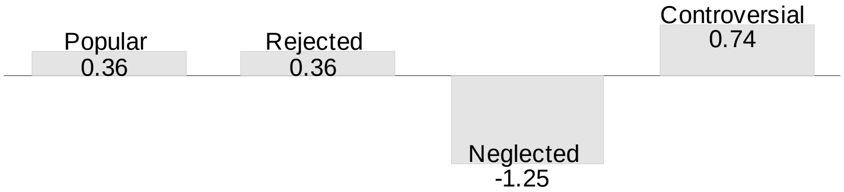


Figure 5: Effect size of varying interaction probability (the average number of deviations a group is from the mean)

Finally, we examined the finding that controversial children tend to display a combination of positive and negative social behaviors (e.g., Coie & Dodge, 1988). We modeled this lack of consistency by varying the standard deviation of the Gaussian distribution in Formula 1. The results, displayed in Figure 6 supported our interpretation that Controversials tend to be highly variable in their behavior. The results also revealed an unexpected effect in which neglected individuals tended to be more reliable in their behavior (lower variability). Note also that making children more variable produces approximately the same distribution pattern as making them interact less frequently. This makes sense since highly variable behaviors would result in more actions that cancel each other out, which would approximate the effect of simply not interacting as much.

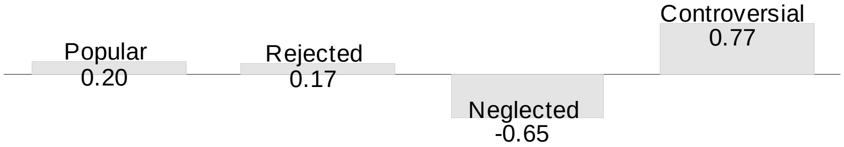


Figure 6: Effect size of varying behavior deviation (the average number of deviations a group is from the mean)

3.4.Dynamic Effects

In the previous section we looked at whether adding biases, in the form of personality or social skills, would produce the expected effects. We found that it did, and that it produced other, unexpected effects as well. The underlying assumption behind biasing the model in this way was the assumption that these individual differences influence popularity. However, we can turn this assumption around and ask if the cognitive dynamics of the model could be generating these popularity effects. That is, could individual differences in cognitive factors underlying the dynamics of peer group formation lead us to display individual differences that appear to be due to our personality or level of social skills.

The first that we looked at was the initial value for a[i,j]. For this simulation some agents started off more predisposed to liking everyone (a high value of a[i,j] for all j), and others more predisposed to disliking everyone (a low value of a[i,j] for all j). This was meant to represent the effect of learning, before entering the group (e.g., family experiences or other peer group experiences). Thus there were no parameterized individual differences; only the stored information at the beginning was different. The idea behind this was to see if the model displayed sensitive dependence on its initial conditions, a common feature in dynamic systems. The results of this simulation are displayed in Figure 7. Note that the results were very similar to the results of the hostile attribution simulation and the social skills simulation. All of these showed an expected association between strong positive biases and being popular, and strong negative biases and being rejected. They also all showed the unexpected association between moderate negative biases and being neglected.

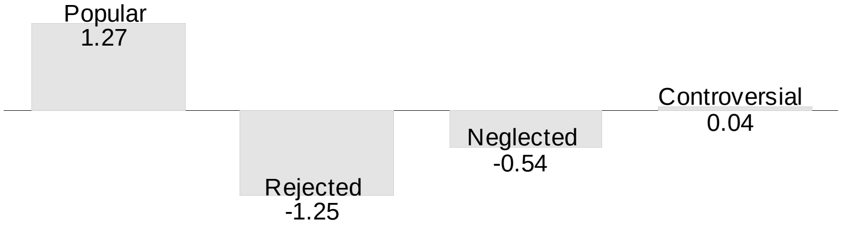


Figure 7: Effect size of varying the initial value of a[i,j] (the average number of deviations a group is from the mean)

We also analyzed the effect of individual differences on r. That is, how does having a different adaptation rate affect one's eventual CDC categorization? This was to explore the effects of individual differences in the only cognitive parameter in the model. The results, displayed in Figure 8, were interesting, in that they were very similar to the results of the interaction probability simulation and the behavior consistency simulation. The most straightforward interpretation of this is that having a high learning rate or having many interactions can have the effect of making one appear to be more variable in behavior.

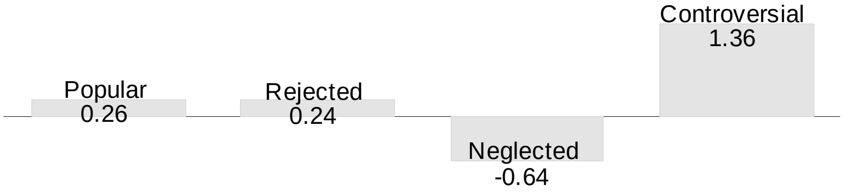


Figure 8: Classification effect size of varying r (the average number of deviations a group is from the mean)

These two graphs represent novel predictions of our model that could be confirmed experimentally. In particular, Figure 8 seems the most amenable to direct confirmation of our model's results. A test of learning rate or speed of adaptation could be applied to various students, along with a CDC questionnaire.

The results on the effects of individual differences can be divided into two categories, which are shown in Figures 9a and 9b. The key point to observe here is that there are only two major effects caused by varying these six different individual traits. In Figure 9a, we see that Popular individuals tend to be those who interpret others actions in a positive light, tend to behave better towards others, and who start with a positive attitude towards others. Rejected individuals tend to be negative in all of these traits. This is not a particularly surprising result. What is surprising is that these same three traits are also grouped together for Neglected individuals and for Controversials. People who are Neglected by their peers tend to be slightly negative in these three traits, and Controversials are the same as Average people in these traits. The other three traits are also closely grouped, as shown in Figure 9b. Interacting more often, having more random deviations in one's actions, and changing one's beliefs quickly in response to others' actions are all traits common in Controversial people. Neglected people, on the other hand, tend to be low in those same three traits. Furthermore, both Rejected and Popular people tend to be moderately above the average.

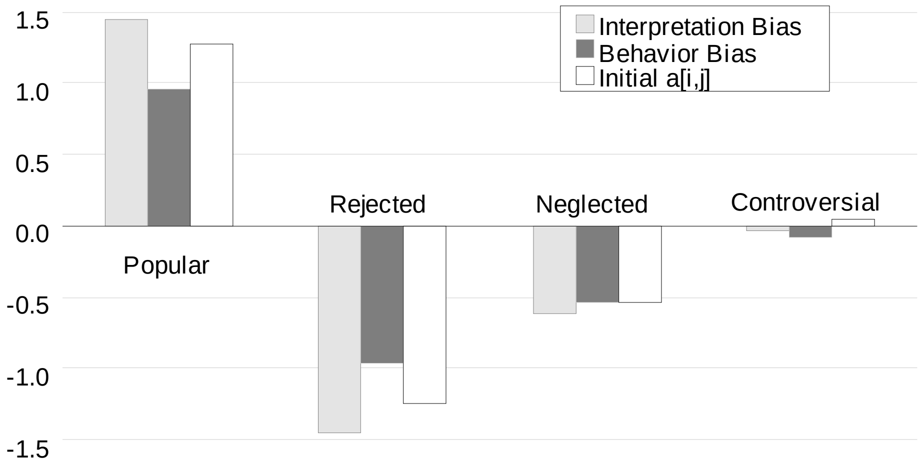


Figure 9a: Three associated individual difference effects

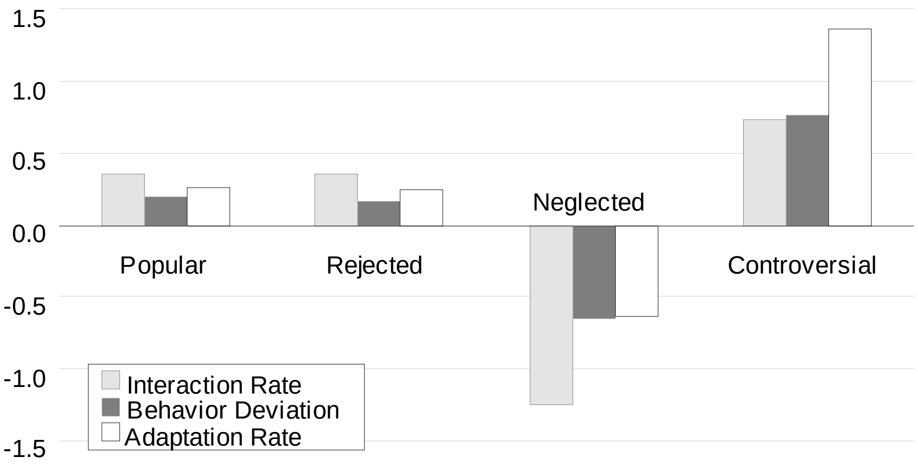


Figure 9b: Three more associated individual difference effects

These results show that individual differences in the cognitive factors (initial memory contents and adaptation rates) underlying the dynamics of peer group formation can produce individual differences that appear to be due to our personality or level of social skills. This result indicates the possibility that personality and social skills may in fact be generated by cognitive factors interacting with the dynamics of peer group formation. This is a very different interpretation from the standard view that personality and social skills are relatively static traits and abilities that individuals possess. We do not wish to claim that static traits and abilities do not affect popularity, especially amongst adults. However, it may be the case that the genesis of these traits and abilities lays in our early experiences with the dynamics of peer group interactions. That is, children learn concrete forms of certain traits and abilities through displaying them – and they initially display them because of the interaction between their cognitive system and the dynamic of peer group interactions.

4. ACT-R Modeling

In the preceding sections of this paper, we used the simplicity principle to produce a computational model that can account for the data on children's popularity. To extend this, we now describe an ACT-R version of the model, thus incorporating the architecture principle. The intent is to show that our simple model can also be realized in a system designed to be similar to the way human cognition actually works, as opposed to an arbitrary mathematics-based model.

However, the ACT-R architecture imposed a major difference. ACT-R makes the strong claim that exact *magnitude* is not something that is directly accessible to cognition. That is, a person might choose to be nice or not-nice, but cannot choose to be 1.7 nice or 2.48 not-nice. While it is possible to approximate magnitudes in ACT-R by using the architecture to construct a discrete magnitude scale, this would not represent the simplest and most direct use of the architecture. Therefore, we restrict the actions for the ACT-R model to being either nice or not-nice, with no magnitude component.

Another issue for the ACT-R implementation was how to keep track of how other people have behaved in the past. The ACT-R, procedural memory mechanism for learning the utility of actions could be used to do this in a way that is similar to our simple model. However, this would involve having separate and distinct procedural rules for every different person, which seems suspect (rather like having different production rules for driving different makes of cars). In contrast, we know that people store information about individuals in declarative memory, so we choose to use the ACT-R declarative memory system as the basis for our model. As we shall show, this resulted in a qualitatively different type of model from our simple model.

The ACT-R declarative memory system does not have the ability to learn utility values over time. However, it can be used to detect sequential dependencies in data generated across time (Lebiere & Wallach, 2001; Lebiere & West, 1999). In addition, this mechanism has been shown to account for both implicit learning and game playing experimental results. Following from this, we re-formulated friendship interactions as a matching game, where you want to predict if someone will behave positively or negatively to you, and then match that prediction with your own actions (i.e., if you expect someone to be positive then you act positively towards them).

The ACT-R friendship model makes these predictions by using sequential dependencies to predict, from their previous *n* interactions, what an individual will do on this interaction. Although ACT-R was not designed to detect sequential dependencies, it turns out that there is a straightforward way to do this within the architecture. After each interaction, a record of what occurred, along with a record of the previous *n* interactions, is stored in the ACT-R declarative memory system as a *chunk*. Each time the same sequence of events is observed it strengthens the activation of that chunk in memory. Thus, chunk activation level reflects the past likelihood of a sequence occurring. When the system attempts to predict what will happen next, it can makes use of its current knowledge of the previous n interactions with this individual to retrieve the matching chunk with the highest activation.

For example, if the opponent's last move was P (where P is positive and N is negative) and the model was set to use information from only the previous move (i.e., n=1; also termed *lag 1*), then there would be two possible matching chunks: PP (where the other person acted positively twice in a row) and PN (where the other person acted positively and then negatively). The activation levels of these chunks depend on the timing of previous situations where these sequential events occurred, according to the standard ACT-R base level learning formula.

$$A_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \varepsilon \qquad \text{(Formula 3)}$$

A_i is the activation of the i^th^ chunk, t_j is the amount of time since seeing the j^th^ example of the chunk, *d* is a decay parameter (always set to 0.5), and ε is random logistic noise (with a standard deviation around 0.3). Neither parameter was adjusted from these default values for our model.

Thus, if PN had the highest activation (i.e. if this particular person had often responded negatively after being positive) then the model would respond negatively to match that expectation. The model would then observe what the other agent actually did and store a record of it. That is, if they actually responded positively, then the chunk indicating the sequence PP would be strengthened.

Importantly, all of these modeling details are generalized from earlier work, and were not designed specifically for this model. That is, our ACT-R model makes use of the established results on sequential dependency detection in the same way that it makes use of the established results on ACT-R modeling in general. Our use of the ACT-R framework allows for the creation of a model that is simple to describe within the language of ACT-R, but at the same time inherits the properties of making use of known psychological and neurological results.

To run the simulations, we followed the same process as with the previous model. We created thirty ACT-R models (one for each individual), and allowed them to interact over time. To do this we used Python ACT-R project (Stewart and West, 2005). Python ACT-R is an implementation of the ACT-R theory in the Python programming language, and is particularly useful for making multi-agent simulations with explicit control over the application of the ACT-R theory.

It should be noted that by making use of the ACT-R architecture, we are also addressing a common criticism made about cognitive modeling: parameter fitting. This is the situation where a sufficiently flexible model can be made to fit any experimental data by adjusting various parameters. Indeed, our initial computational model only gives results which match the experimental results when the adaptation rate parameter is between 0.05 and 0.2. Fortunately, this is a very wide range, which has been identified as an important criterion for a model to be persuasive (Roberts & Pashler, 2000). However, a model is even more convincing *if there are no free parameters*. To achieve this in our ACT-R model, we left all of the ACT-R parameter settings at their default values, which are the settings that have been found to work best across numerous different simulations. They are also the settings that were found to work best for the game playing model (Lebiere & West, 1999), on which this model is based. Also, the model code was not customized in any way to fit the data. The friendship model is directly based on the game model, which was directly based on the basic implicit learning model of ACT-R. Thus, our claim is that thinking about friendships is handled by the same system used for game playing and implicit learning, meaning that these previous models completely constrain our current model. The ACT-R model can therefore be seen as an absolute prediction, with no possible room for parameter tweaking.

As before, the key dataset for comparison is the stability of category memberships. This was measured in an identical manner as with our computational model. Figure 10 shows our results compared to the expected real-world values.

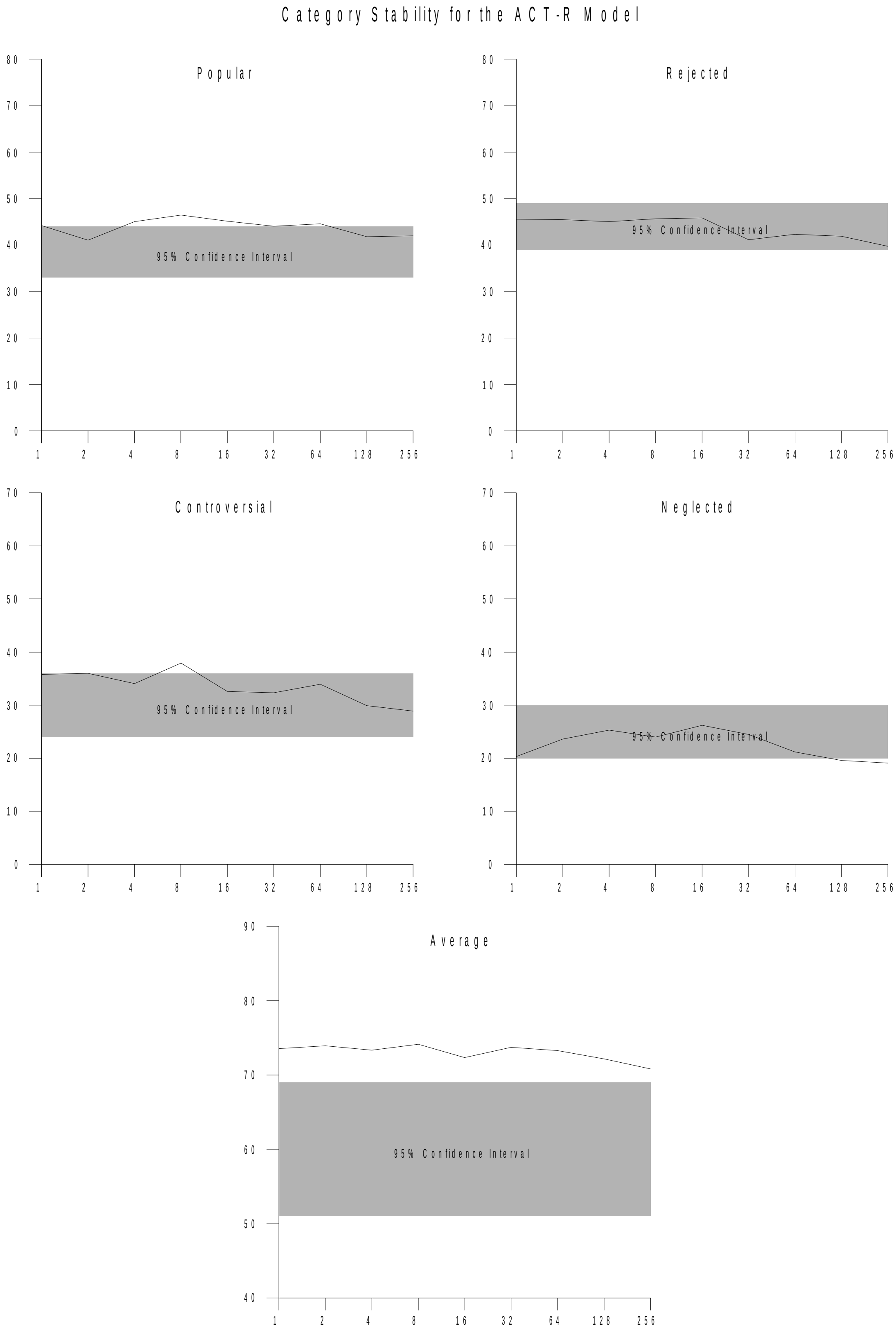


Figure 10: Stability results from the ACT-R model, compared to the 95% confidence interval for the real-world data.

From these graphs we can see that the ACT-R model matches extremely well with the human results; in fact, it matches as well as the best of the earlier models. It is worth repeating that *no parameters were modified* to achieve this.

Indeed, in the measurements that it doesn't quite match, it disagrees with the human results in the same way as the simple model (the stability for Average is a bit high, and for Neglected a touch low). This is remarkable, given that these are two completely different models of the same process.

This result deserves more attention. We believe that this is further indication that it is the dynamics of the situation which is the major contributor to the eventual categorization of individuals. Given that these two starkly different models give similar predictions, it is likely that *no matter how individuals actually do choose to behave towards others*, we should get similar CDC patterns, as long as individuals do behave towards others in a manner that's based on how others have behaved towards them (remembering that our initial baseline random-behavior model did not show this pattern).

4.1. Individual Differences

The ACT-R model was more constrained than our other model, since experimental research with ACT-R has uncovered specific parameter settings that tend to lead to models which predict human behavior well. However, it is still possible to introduce individual differences within the model, and we can investigate their effects in the same manner as before.

The first such individual difference is the *Interaction Rate*. This is identical to the interaction rate variable in the previous model; the chance of two individuals interacting is the product of their interaction rates. This value is always positive and can vary both above and below 1 (a probability above 1 is treated as a 100% chance of interaction).

The next individual difference is *Choice Rigidity*. This is an indication of how strict an individual is at acting based on their predictions. An individual with a low choice rigidity may predict someone will behave positively even if the relevant positive chunk has slightly lower activation than the negative one (or vice versa). This is implemented via the ACT-R declarative memory instantaneous noise parameter, which adjusts the variation in the logistic noise (ε) in the activation equation (Formula 3). A high choice rigidity is a low variation in noise, while a low rigidity is a high amount of noise.

For the *Initial Memory Bias*, we adjusted the initial activation of the chunks indicating positive or negative behaviour. An individual with a high positive initial memory bias would have chunks such as NP and PP increased in strength, while a very negative initial memory bias increases the strength of the NN and NP chunks. This was accomplished by adding a fixed value to the activation equation for each chunk.

The *Remembering Bias* indicates a likelihood of treating a single experience as multiple experiences. A positive value indicates that positive events are more likely to be stored in declarative memory twice. This translates to two entries in the sum in Equation 1. This approach is equivalent to a technique from [ref] where important events in a game are *rehearsed*, giving them greater activation strength. It should be noted that ACT-R makes the strong claim that models must perform a whole number of repetitions, and so there is no mechanism for increasing the number of presentations of activation of a chunk by a fractional amount.

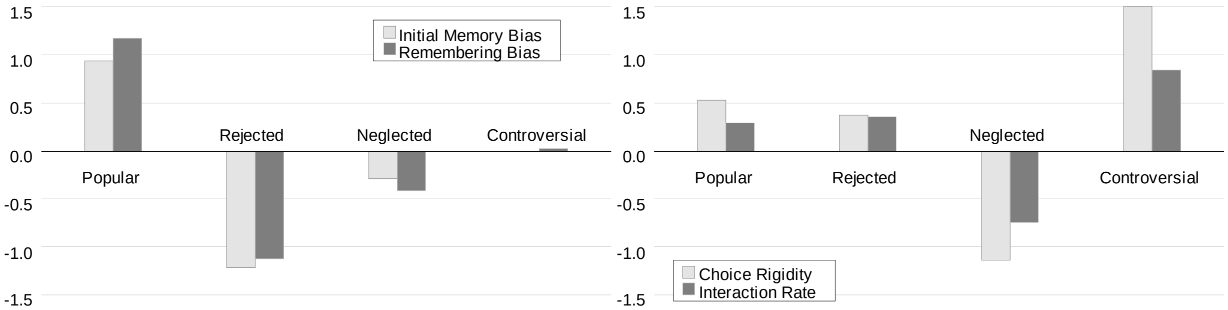


Figure 11: Effect sizes for four individual differences in the ACT-R model.

Figure 11 shows the effect size of these four changes, separated into two groups. Comparing these results to Figure 9, we can see that these are the *exact same effects as with the previous model*. That is, the ACT-R model also shows the same patterns of effects as the earlier model. This is true even though the ACT-R model has an extremely different implementation from that earlier model.

We believe that this indicates that this pattern of the effects of individual difference is something fundamental to this sort of model of peer interaction. Furthermore, if this pattern does arise in many different models, then it is likely to be found in real human interaction as well.

5. Conclusions

There are a number of intriguing results from this research. First and foremost, we have two rather different process-based computational models of how the group dynamics of friendship change over time. Both of these models produce data which is statistically indistinguishable from real human data, in terms of the standard CDC categorization system. Both of these models achieve this close match to the known human data without recourse to extensive parameter fitting. The first model relies on only a single parameter, which is set anywhere in a range consistent with other adaption rate parameters in other learning models. The second model uses *no free parameters at all*, as everything is set to the standard ACT-R default values, which have been successfully used in a wide variety of other cognitive models.

This does not, of course, mean that these models detail the exact inner working of real human friendship formation. Instead, the intended interpretation is that these models accurately capture *those aspects of peer group formation that are revealed by CDC sociometric measures.* That is, to the extent that the CDC metric is useful, these models are also appropriate. To develop a richer understanding of the process, we need not only to develop new models, but also to develop more detailed measurements.

Furthermore, when we introduced individual variability, both models produced data consistent with standard psychological results. Importantly, *none of these results were considered when creating the models*. That is, the models were designed solely to account for known CDC sociometric classification results, and yet also exhibited the Hostile Attribution Bias (Figure 3), standard effects of variations in social skill levels (Figure 4), and the shyness effect in Neglected children (Figure 5). Indeed, these can be seen as confirmed predictions of the models.

Along with these predictions, the models also exhibited other effects which have not been extensively studied in children. Low variability in behavior tends to, in our models, lead individuals to be classified as Neglected (Figure 6). A similar effect is seen in those with a slight tendency to interpret others' behaviors negatively (Figure 3), or to behave slightly negatively (Figure 4). Both Popular and Rejected individuals also have a tendency to interact slightly more often, and to have a higher variability of behavior (Figures 5 and 6). Most interestingly, we also found that Controversial individuals tend to adapt very quickly to others' actions, while Neglected individuals are much slower (Figure 8). Both Popular and Rejected individuals adapt slightly faster than average. Equivalent effects are seen in both models (Figure 11).

These observations are not, in and of themselves, proof that such effects would also be observed in real children. However, we believe that given the accuracy of the other confirmed predictions, it is at least worth further examining these possibilities. Continuing this research in this direction serves two purposes: first, it identifies potentially relevant aspects of peer group relations that may not have been considered otherwise; and second, it allows us to further test the accuracy of our models.

It is also worth noting that we have identified two characteristic patterns caused by individual differences. Each of the 10 different ways we allowed individual to differ ended up causing one of only two different classification distributions (Figures 9 and 11). For five of these differences, the observed result was that Popular individuals tended to score high on that difference, and Rejected individuals scored low. By itself, this is not a particularly surprising result, as there is an inherent opposition in how these are measured. However, *in each case*, that same individual difference would consistently be slightly lower than average in the Neglected individuals, and Controversials were exactly at average. This is a surprising pattern, since these individual differences are themselves also very different. The identical effect is seen whether it is due to how that person interprets others, behaves, is initially biased, or recalls past events. Finding this underlying similarity is, we believe, an important result of this research.

This is especially true given that the other five individual differences considered also showed a consistent and coherent pattern. When varying how often individuals interacted, how quickly they changed their beliefs, how widely the varied their own behavior, or how rigid they were in their choices, the models always produced the same results: Controversials were high on every measurement, Neglected individuals were low, and both Popular and Rejected classifications showed a moderately above-average level.

This also reveals an interesting asymmetry between the two patterns. Individual traits which tended to lead to someone being Rejected would always be seen at a moderate level in Neglected individuals. However, those which tended to be seen strongly in Neglected individuals always had *the opposite effect* (to a moderate degree) for those classified as Rejected. Furthermore, traits seen strongly in Controversials were always seen to smaller degree in both Popular and Rejected individuals, but traits which lead to either of those classifications *would consistently have no effect* on being classified as Controversial. This pattern gives us a new way of looking at the how these internal individual factors interact to produce the overall observable sociometric classification results. Furthermore, the fact that these patterns consistently appear for two highly different models leads us to believe that these predictions may hold for almost any reasonable model of this situation, and thus we have identified an important fundamental aspect of real-world peer group formation.

6. References

Anderson, J. R. & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.

Baumeister, R. F., Campbell, J.D., Krueger, J.I., & Vohs, K. D. (2003). Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?, Psychological Science in the Public Interest, 4(1), 1–44.

Bouchaud, J-P. & Mézard, M. (2000). Wealth condensation in a simple model of economy, *Physica A*, *282*, 536.

Cassidy, J., Scolton, K.L., Kirsh, S.J., & Parke, R.D. (1996). Attachment and representations of peer relationships. *Developmental Psychology, 32*, 892-904.

Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science? Trends in Cognitive Sciences, 7(1), 19-22.

Cillessen, A., Bukowski, W. (2000). Conceptualizing and Measuring Peer Acceptance and Rejection. *New Directions for Child and Adolescent Development, 88*, 3-10.

Cillessen, A., Bukowski, W., & Haselager, G. (2000). Stability of Sociometric Categories. *New Directions for Child and Adolescent Development, 88*, 75-93.

Cillessen, A. & Brink, P. W. M. (1991). Vaststelling van relaties met leeftijdgenoten [Assessment of children’s peer relationships]. Pedagogische Studiën, 68, 1–14.

Coie, J.D. & Dodge, K.A., (1988). Multiple sources of data on social behavior and social status. *Child Development, 59*, 815-829.

Coie, J.D., Dodge, K.A., and Coppotelli, H. (1982). *Dimensions and Types of Social Status: A Cross-Age Perspective*. Developmental Psychology, 18, 557-570.

Coie, J.D., Dodge, K.A., & Kupersmidt, J.B. (1990). Peer group behavior and social status. In S.R. Asher & J.D. Coie (Eds.), *Peer rejection in childhood*. New York: Cambridge University Press.

Cohn, D.A. (1990). Child-mother attachment of six-year-olds and social competence at school. *Child Development, 61*, 152-162.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101.

Dodge, K.A., Coie, J.D., & Brakke, N.P. (1982). Behavioral patterns of socially rejected and neglected preadolescents: The roles of social approach and aggression. *Journal of Abnormal Child Psychology,* 10, 389-409.

Dodge, K. A., Lansford, J., Burks, V., Bates, J.E., Pettit, G., Fontaine, R., & Price J.M. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374-393

Elicker, J., Englund, M., & Sroufe, L.A. (1992). Predicting peer competence and peer relationships in childhood from early parent-child relationships. In R. Parke & G. Ladd (Eds.), *Family-peer relationships: Models of linkage*. Hillsdale, NJ: Erlbaum.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology.* 83(6)*,* 1420-1443.

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating Dynamical Features of Escape Panic. *Nature 407*, 487-490.

Lebiere, C. & Wallach, D. (2001). Sequence Learning in the ACT-R Cognitive Architecture: Empirical Analysis of a Hybrid Model. In R. Sun & C. L. Gilles (Eds.). Sequence Learning: Paradigms, Algorithms, and Applications. Berlin: Spinger Lecture Notes in Computer Science.

Lebiere, C., & West, R. L. (1999). A dynamic ACT-R model of simple games. In Proceedings of the Twenty-first Conference of the Cognitive Science Society, pp. 296-301. Mahwah, NJ: Erlbaum.

Maassen, G. H., van Boxtel, H. W., & Goossens, F. A. (2005). Reliability of nominations and two-dimensional rating scale methods for sociometric status determination, Journal of Applied Developmental Psychology. 26(1), 51-68.

Moreno, J. L. (1934). *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. Washington, D. C.: Nervous and Mental Disease Publishing Co.

Newcomb, A.F., & Bukowski, W.M. (1983). *Social Impact and Social Preference as Determinants of Children's Peer Group Status*. Developmental Psychology, 19, 856-867.

Newcomb, A.F., Bukowski, W.M., & Pattee, L. (1993). Children's peer relations: A meta-analyic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin,* 113, 99-128.

Newell, A. (1990). Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts.

Parkhurst, J.T., & Hopmeyer, A. (1998). Sociometric popularity and peer-perceived popularity: Two distinct dimensions of peer status. *Journal of Early Adolescence,* 18, 125-144.

Rescorla, R.A, & Wagner, A.R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II, A.H. Black and W.F. Prokasy, Eds. 64-99.

Roberts, S. & Pashler, H. (2000). "How persuasive is a good fit? A comment on theory testing". Psychological Review. 107 (2), 358-367.

Rubin, K. H. & Mills, R. S. (1988). The Many Faces of Social Isolation in Childhood. Journal of Consulting and Clinical Psychology, 56, 6, 916-924.

Rubin, K.H., Hymel, S., & Mills, R.S. (1989). Sociability and social withdrawal in childhood: Stability and outcomes. *Journal of Personality, 57*, 237-255.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533-536.

Stewart, T. C. & West, R. L. (2005) Python ACT-R: A New Implementation and a New Syntax. 12th Annual ACT-R Workshop.

Sun, R. (2001). Cognitive science meets multi-agent systems: a prolegomenon. *Philosophical Psychology*, 14(1), 5-28.

Sun, R, ed. (2006) Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation. Cambridge University Press, New York.

Terry, R. (2000). Recent Advances in Measurement Theory and the Use of Sociometric Techniques. *New Directions for Child and Adolescent Development,* 88, 27-53.

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.