Social Categorization Through the Lens of Connectionist Modeling

André Klapper (a.klapper@psych.ru.nl)
Radboud University Nijmegen, Behavioural Science Institute
Postbus 9104, 6500 HE Nijmegen, Netherlands

Iris van Rooij (i.vanrooij@donders.ru.nl)
Radboud University Nijmegen, Donders Institute for Brain, Cognition, and Behaviour
Postbus 9104, 6500 HE Nijmegen, Netherlands

Ron Dotsch (r.dotsch@uu.nl)
Utrecht University, Social and Organizational Psychology
P.O. Box 80.140, 3508 TC Utrecht, Netherlands

Daniël Wigboldus (d.wigboldus@psych.ru.nl)
Radboud University Nijmegen, Behavioural Science Institute
P.O. Box 9104, 6500 HE Nijmegen, Netherlands

Keywords: Categorization; Connectionism; Person Perception; Stereotyping; Cognitive Modeling; Computational Explanation

Social psychology does not yet have a strong cognitive modeling tradition. This is not for lack of cognitive modeling tools that are relevant and useful for modeling social psychological phenomena. For instance, several researchers have successfully demonstrated how connectionist modeling techniques can be used to build computational explanations of key phenomena of interest to social psychologists, such as stereotyping, prejudice and priming (Kunda & Thagard, 1996; Schröder & Thagard, 2014). In this project we contribute to this important development by addressing a major obstacle to the progression of connectionist modeling in social psychology: That is, how can we reconcile the intuitive concepts that figure in the verbal explanations that pervade social psychological theories with formal properties and processes in connectionist models? We illustrate a systematic way of addressing this question by considering the theoretical concept of 'social categories', which plays a central role in social psychological theories. Using computer simulation, we show that if social categories are defined as 'excluders' in connectionist models then key social psychological phenomena can be replicated, while maintaining a clear link with the intuitive concept of social categories. We discuss the broader implications of our simulation results for both social psychology and cognitive modeling.

Existing Person Perception Models

Social categorization theories belong to the most prominent verbal theories in the area of person perception (Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000). A general claim in these theories is that stereotyping and prejudice are the result of the natural tendency of people to categorize perceived people. A central assumption in these theories is that people construed other people based on two types of mental representations: social categories (e.g. gender, nationality, or occupation) and attributes (e.g. personality traits or physical features). It is further assumed that if a person categorizes another person then that triggers a set of (implicit) beliefs about the categorized person in the perceiver. This set of (implicit) beliefs is referred to as the *stereotype* of the category. In contrast, if attributes are assigned to the person, no such (or much fewer) beliefs are triggered. Hence, in social categorization theories, social categories are the main cause of stereotyping.

More recently, (localist) connectionist models of person perception have been proposed, which explain stereotyping and prejudice by the spread of activation between mental representations via associative links (Freeman & Ambady, 2011; Kunda & Thagard, 1996). Every mental representation in these models is associated with other mental representations, which means that every mental representation (and thus not only a particular subset) can trigger associated beliefs, in principle. Perhaps for this reason, Kunda and Thagard (1996) have presented their connectionist model as an (competing) alternative to social categorization theories. In contrast, Freeman and Ambady (2011) proposed that the general process of social categorization may be implemented by a connectionist process. These conflicting perspectives illustrate that the relationship between connectionism and social categorization has remained relatively unclear. If, and how, the different perspectives can be reconciled is thus an important open problem.

What is a Social Category?

A major obstacle to unifying social categorization and connectionist models is that the verbal term 'social category' leaves too much room for interpretation. As a first step towards an integrative model, we disentangle the most prominent interpretations. We argue that most interpretations are either in conflict with major assumptions of social categorization theories or with empirical evidence. Based on this, we argue for an interpretation in which social categories can roughly be described as '*excluders*': that is, mental representations that strongly exclude some other mental representation. This interpretation can be implemented in a connectionist model by giving those mental representations that are conceived of as categories strong inhibitory connections that prevent their co-activation (see Fig. 1).

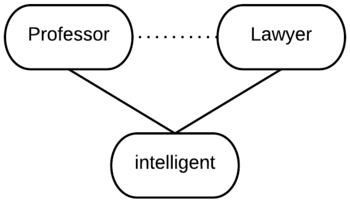

Figure 1: Illustration of a connectionist network of 'social categories' and 'attributes'. Excitatory connections are denoted by solid lines and inhibitory connections by doted lines. Within this network, *Professor* and *Lawyer* are social categories because they are connected by a (strong) inhibitory connection. In contrast, *intelligent* is an attribute because it does not have a (strong) inhibitory connection.

An Integrative Model

In our poster, we will demonstrate how social categories under our connectionist interpretation give rise to key phenomena that have been attributed to (social) categorization. Specifically, we will present simulation results that show how our connectionist interpretation of social categories gives rise to stereotyping in a way that is consistent with the general assumptions of social categorization theories.

Andersen and Klatzky (1987) provided empirical evidence that people can infer more varied characteristics about a person when provided with a category label (e.g. *professor*) compared to a trait label (e.g. *intelligent*). We replicate these results in our connectionist simulation in which activating a category (under our interpretation of categories) by external input leads to the activation of more other mental representations compared to a situation in which an attribute is activated by external input. In other words, category activation triggers more (stereotypical) beliefs than attribute activation, which does not only explain the results by Andersen and Klatzky but also conceptually replicates the category-attribute distinction in social categorization theories.

Furthermore, we show how inhibitory associations generate the general phenomenon attributed to categorization that the subjective similarities of people within categories and is decreased and the subjective similarities of people between categories is increased. This gives rise to the well-replicated phenomenon that discrimination performance is highest for stimuli that are separated by a category boundary (Goldstone & Hendrickson, 2009).

Conclusions

We replicate key social psychological phenomena that have been attributed to social categorization processes in a formal connectionist model. In addition, we provide a clear mapping of the verbal terms of social categorization theories (in particular, the terms ‘categories’ and ‘attributes’) onto formal connectionist properties. This unifies social categorization and connectionist models of person perception. Moreover, our approach demonstrates a possible way to reconcile the verbal approach taken in social categorization theories with the formal approach taken in connectionist models. That is, while the connectionist model of the social categorization process provides formal precision, the intuitive concepts ‘categories’ and ‘attributes’ provide useful verbal heuristics that summarize the functional behavior of these different mental representations in (connectionist models of) person perception. This creates a bridge between the verbal theorizing in social psychology and the formal modeling in the connectionist literature, which makes it possible for the two research areas to inform each other more in the future.

Acknowledgments

This research was supported by an NWO grant 464-11-036 awarded to Ron Dotsch.

References

Andersen, S. M., & Klatzky, R. L. (1987). Traits and social stereotypes: Levels of categorization in person perception. *Journal of Personality and Social Psychology*, *53*(2), 235–246.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression for- mation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. Zanna (Ed.), *Advances in experimental socialpsychology*. San Diego, CA: Academic Press

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–79.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 69-78.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors : A parallel-constraint-satisfaction theory, *Psychological Review*, *103*(2), 284–308.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Schröder, T, & Thagard, P. (2014). Priming: Constraint satisfaction and interactive competition. *Social Cognition, 32*, 152-167.