

MULTIPLE PERSON INFERENCES: A VIEW OF A CONNECTIONIST INTEGRATION

FRANK VAN OVERWALLE

Vrije Universiteit Brussel.

Pleinlaan 2,

B-1050 Brussel, Belgium

E-mail: Frank.VanOverwalle@vub.ac.be

This paper provides a connectionist account of the processes underlying the multiple inference model of person impression formation proposed by Reeder, Kumar, Hesson-McInnis and Trafimow (2002). First, in a replication and extension of one of their main studies, I found evidence for discounting of trait inferences when facilitating situational forces were present consistent with earlier causality-based theories, while at the same time I replicated the lack of discounting in moral inferences as documented and predicted by Reeder et al. (2002). Second, to provide an account of how these different and sometimes contradictory inferences are formed and integrated in a coherent person impression, I present a recurrent network model that automatically integrates these inferences, resulting in a pattern that closely reproduces the observed data.

1. Introduction

Although person impression research is replete with studies on how context and behavior interact to determine one's impression about a person, in the past, little attention has been given to other inferences that people may make in the process. In a recent study, Reeder, Kumar, Hesson-McInnis and Trafimow (2002) proposed a *multiple inference model* to account for the many inferences that are made and used to arrive at a person impression. They documented that besides trait and causal inferences, perceivers routinely infer much more about the actor, such as his or her motives for engaging in a given behavior and its social implications. Although it is widely accepted that perceivers take note of an actor's goals, Reeder et al.'s important insight is that these inferences impact on the impression formation process. For instance, when an actor engages in aggressive behavior, we may infer not only to what extent he or she is an aggressive person, but also which reasons or motives behind the aggressive act may have compelled the actor to behave the way he or she did. By taking this information into account, Reeder et al. argued, we may end up with a completely different impression about the person. This is interesting, because it can potentially explain why people sometimes fall prey to attribution biases, like the fundamental attribution error. This bias refers to the well-known phenomenon that in explaining an actor's behavior, observers often do not sufficiently take into account the facilitative contextual forces and fail to

use this information to discount the contribution of the actor. Perhaps, motives may explain in part why this bias occurs.

The aim of the present paper is to replicate and extend the crucial finding of Reeder et al. (2002) that perceivers apparently fail to apply discounting when inferring a behavior-implying trait due to additional inferred motives. In addition, because Reeder et al. (2002) did not provide a formalization of how multiple inferences are derived, I propose a connectionist account of this process. This is part of an ongoing attempt to provide more precise computational implementations of diverse phenomena and processes in social cognition. Inspired by the ever-increasing success of connectionist models in cognitive psychology, a number of social psychologists developed connectionist models of causal attribution (Van Overwalle, 1998, Read & Montoya, 1999), cognitive dissonance (Shultz & Lepper, 1996; Van Overwalle & Jordens, 2002), group impression formation and change (Kashima, Woolcock & Kashima, 2000; Van Rooy, Van Overwalle, Vanhoomissen, Labiouse & French, 2003) and person impression formation (Kashima & Kerekes, 1994; Smith & DeCoster, 1998; Van Overwalle & Labiouse, 2004).

2. Multiple Inferences

To illustrate the multiple inference perspective, consider the following scenario used in Reeder et al.'s (2002) research. John, a participant in a psychological experiment, is told that he has the choice to either reward another research participant or punish the other participant, by giving an extra dollar or delivering a shock. John is further told that the other participant is facing a similar decision about whether he would reward or shock John, and that the other participant would be the first to choose. In one situation, the other participant decided to deliver the shock, and thus encouraged an aggressive response from John. In the other situation, the other participant decided to reward John, and thereby discouraged an aggressive response from John. In both situations, however, John decided to deliver a shock to the other participant. In which situation would we consider John the most moral or the most selfish character?

Reeder et al.'s (2002) hypothesis was that inferences about morality are based on the perceived motives of the target. In the above aggression encouraging scenario where the other participant punished first, the perceived motive may include revenge, so that John's aggression in response to this sort of provocation seems quite natural and legitimate. Reeder et al. (2002) argued that "motives of this sort may be viewed as relatively positive (or less negative)" and "it follows, therefore, that perceivers' inferences about morality should be more positive" (p. 792). In contrast, other aggressive situations may revolve around a selfish desire for profit. For instance, imagine the same

scenario but now no mention is made of the other participant's choice, but rather that the experimenter either encouraged aggression by offering John a \$5 incentive for delivering the shock or discouraged aggression by offering the same incentive for rewarding the other participant. In this aggression encouraging situation, Reeder et al. (2002) argued, "the potential motives underlying instrumental aggression converge on selfish desires for reward. Perceivers may react negatively to such motives" (p. 792).

Consistent with their hypothesis, across four experiments, Reeder et al. (2002) found a strong decrease of morality when the aggressor acted out of selfishness as opposed to revenge. Thus, when John reacted to the other participant's punishment that encouraged aggressiveness (*reactivity condition*), he was seen as more moral and less selfish than when the experimenter provided financial incentives that encouraged aggression (*instrumentality condition*). These findings indicate that the motives underlying someone's behaviors seem to contribute strongly to inferences of morality. Hence, Reeder et al. (2002) concluded that "inferences about the morality of an aggressor are based more on the perceived motives of the target than on the presence of facilitating situational forces" (p. 789).

3. Discounting and Motives

An important implication of Reeder et al.'s (2002) findings is that they challenge general causality-based models of dispositional inference such as those of Kelley (1971) and Gilbert (1989), in that perceivers apparently failed to apply Kelley's discounting principle. According to the discounting principle, perceivers consider whether the behavior appears to have been caused by the actor's disposition or by situational forces, and assume that both causes operate in a hydraulic fashion. Hence, when the situation encouraged aggression, perceivers should subtract out the effect of the situation, with the result that only a minimal amount of aggression is attributed to the person. Conversely, when the situation discourages aggression, perceivers should not subtract out the effect of the situation, and attribute a great amount of aggression to the person.

In contrast to these assumptions, the results of Reeder et al. showed that although perceivers discounted selfishness in the revenge (reactive aggression) condition, they did not discount selfishness when they were paid for the aggressive behavior (instrumental aggression condition). Reeder et al. argued that this was so because the motive of revenge was legitimate (and not so negative), whereas the motive of selfish gain was not legitimate and is consistent with low morality. Hence they concluded that the "perceiver's use of the discounting principle depended on the type of situational force that was operating" (p. 799).

4. Measuring Traits and Motives: A Replication

Although it is possible that discounting is not applied when making morality judgments, the crucial question is whether perceivers also fail to apply discounting when inferring the correspondent (i.e., behavior-implying) trait? A limitation of most studies by Reeder et al. (2002) is that they did not measure inferences about the correspondent trait aggressiveness. This makes it difficult to compare traits against motives during the same inference process. While a correspondent trait aggressiveness is very close to the actor's aggressive behavior in semantic terms, inferred motives and morality may have greater conceptual breadth and may be more distant to the aggressive behavior. This leaves open the possibility that situational forces are not immediately subtracted out of a moral judgment, as they constitute an essential part in the evaluation and definition of a moral act. Thus, while situational discounting may be absent in judgments of motives and morality, in attributing traits of a person, some discounting of facilitating situational pressures may still take place. In order to verify this hypothesis, I replicated the second study of Reeder et al. (2002) and extended it with a correspondent trait measure.

4.1. Method

Participants were 88 male and female students from the Vrije Universiteit Brussel, Belgium. I used the same materials (translated to Dutch) from Study 2 by Reeder et al. (2002) which depicted John's reactions in the psychology experiment as described earlier in the introduction. To recall, in the *reactivity scenario*, John received a shock from another participant (aggression encouraged) or did not received a shock from another participant (aggression discouraged). The *instrumentality scenario* was similar, except that no mention was made of the other participant's choice. Rather, the experimenter said that he would financially reward John for shocking the other participant (aggression encouraged) or for not shocking the participant (aggression discouraged). All conditions ended by John shocking the other participant.

Immediately following the description of the situation and behavior, the participants rated John on several scales taken from Study 2, except when noted otherwise. Trait aggressiveness was measured with the following item (from Study 4): "How aggressive is John in general, in his everyday interactions" (1 = *not aggressive*; 10 = *very aggressive*). Morality was measured by an item dealing with John's selfishness: "How selfish is John (1 = *not at all selfish*; 10 = *very selfish*) and an item dealing with morality "How moral is John" (1 = *not at all moral*; 10 = *very moral*). In addition, motives were measured by an item dealing with John's motivation to earn money: "To what extent is John motivated to earn money" (1 = *not at all motivated*; 10 = *very motivated*), and John's motivation to revenge himself by an item (from the coding of the open

responses in Study 2): “To what extent is John motivated to revenge himself” (1 = *not at all motivated*; 10 = *very motivated*). Two items served as manipulation checks and asked participants to rate the strength of the situational forces related to provocation: “Did the other participant do anything harmful or inappropriate to John?” (1 = *not at all*; 10 = *very much*), and related to reward: “Could John gain something by hurting the other player? (1 = *not at all*; 10 = *very much*).

4.2. Results

The results showed that the participants discounted *trait aggressiveness* in situations that encouraged aggression in line with causality-based theories, whereas their *moral judgments* were determined by personal motives as Reeder et al. (2002) predicted (see Table 1). An analysis of variance on *trait aggressiveness* revealed only a main effect of situation, $F(1, 84) = 17.06, p < .0001$. The results showed lower ratings of aggressiveness when the situation encouraged aggression than when the situation discouraged aggression regardless of type of scenario.

Table 1: Perceptions of an Aggressive Target Person as a function of Scenario and Situation (Replication Study)

	Reactivity		Instrumentality	
	Aggression encouraged	Aggression discouraged	Aggression encouraged	Aggression discouraged
Trait Aggressiveness	4.78 _a	6.32 _b	4.95 _a	6.45 _b
Morality	4.77 _a	4.09 _a	4.05 _a	3.68 _a
Selfishness	5.59 _b	5.73 _b	6.82 _c	4.18 _a
Motivation to Earn	5.05 _b	4.41 _b	7.91 _b	2.05 _a
Motivation to Revenge	7.50 _c	4.41 _a	4.50 _a	6.00 _b

Note. All cells $N = 22$, except for *Motive to Earn* which contained one missing response. Means in a row with a different subscript differ significantly from each other (Newman-Keuls tests with $p < .05$)

In contrast, replicating Study 2 of Reeder et al. (2002), the predicted interaction between type of scenario and situation was found for ratings of *selfishness*, *motivation to revenge* and *motivation to earn money*, $F_s(1, 84) = 9.49\text{--}49.06, p_s < .01$. No effects were found for ratings of *morality*. For the reactivity scenario, Newman-Keuls tests showed that the motivation to revenge was stronger when the situation encouraged aggression (when the other participant shocked John first) than when the situation discouraged aggression (when the other participant did not shock John), $p < .001$. Conversely, in the

instrumentality scenario, tests showed that ratings of selfishness and motivation to earn money were stronger, and that the motivation to revenge was weaker when the situation encouraged aggression (when the experimenter rewarded shocking the other participant) than when the situation discouraged aggression (when the experimenter rewarded not shocking), $ps < .01$.

Taken together, the results of the aggressiveness ratings are consistent with causality-based theories. We saw discounting of aggressiveness ratings whenever external situational forces encouraged aggressive behavior, irrespective of the type of scenario. In contrast, the selfishness ratings are consistent with Reeder et al.'s (2002) suggestion that participants tend to infer underlying motives and morality that are very different in the two scenarios. When the situation encourages aggression, the actor's aggressive behavior is mainly perceived as being driven by a legitimate motive to revenge in the reactivity scenario, and by a selfish motive to earn money in the instrumentality scenario. The results did not replicate Reeder et al. (2002) morality ratings, presumably because this type of abstract inference is somewhat more culturally dependent and difficult to translate properly with its original (American) meaning preserved.

5. Simulation with a Recurrent Network

The results of the replication study showed that discounting took place for aggressiveness trait inferences in all conditions that encouraged aggression, but not for moral inferences. Thus, inferred traits did not fall prey to the fundamental attribution error, whereas inferred morality did. How can these divergent inferences be explained? Reeder et al. (2002) proposed that an inferred motive tends to be evaluated either positively (or justified) or negatively (or unjustified), and that these evaluative reactions are reconciled with their trait inferences. Thus, inferences of motives and traits are based on their evaluative consistency. However, the authors did not specify the process by which this integration takes place.

The aim of this section is to demonstrate how a connectionist approach may provide an answer to the seeming paradox of divergent motives and traits. I applied the auto-associator network with the delta learning algorithm (McClelland & Rumelhart, 1985) used also by Smith and DeCoster (1998) and Van Overwalle and Labiouse (2004) in their simulations of classic findings in person impression formation. My replication study of Reeder et al. (2002) was modeled using a network architecture consisting of an actor node connected to a trait-implying behavior node (i.e., shock), and additional nodes that reflected situational forces or affordances, including the opportunity to gain money and the other participant (see Figure 1). I also assume that perceivers implicitly categorize the behaviors and motives as socially good (prosocial) or bad

(asocial). Although these implicit categorizations are not essential for the present simulation as they are all asocial, they were included because Reeder proposed evaluative consistency as an essential part of the integration of trait and moral inferences. Simulations of more recent studies by Reeder and colleagues (Reeder, Vonk, Ronk & Ham, 2003) not reported here, show more compellingly that person inferences may depend on these implicit evaluative categorizations.

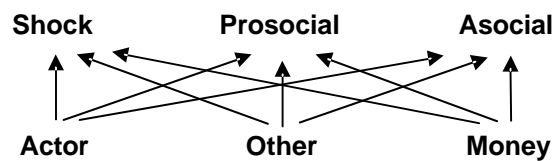


Figure 1. Architecture of the network. Other = Other Participant. Only the most important upward connections are drawn, but all downward and all lateral (in two directions) connections were also included in the simulation.

5.1. Method

I simulated each experimental condition separately. To reflect the idea that perceivers take some time to analyze the information provided, in each condition, I provided two learning trials. Table 2 lists the hypothesized learning history (see also Van Overwalle & Labiouse, 2004).

Table 2: Hypothetical Learning History and Measures of Inferences as a function of Scenario and Situation

	Nodes					
	Actor	Shock	Other	Money	Prosocial	Asocial
	Conditions					
Reactivity / Encouraged	1	1	1	0	0	1
Reactivity / Discouraged	1	1	0	0	0	1
Instrument. / Encouraged	1	1	0	1	0	1
Instrument. / Discouraged	1	1	0	-1	0	1
	Measures					
Aggression	1	?	0	0	-?	?
Selfishness	1	0	0	?	-?	?
Motive to Gain	1	0	0	?	0	0
Motive to Revenge	1	0	?	0	0	0

Note. Representation of the simulation based on the experimental design of Reeder et al. (2002, Study 2); Other = Other Participant; Money = Gain of Money. Cell entries denote external activation to which noise was added that was randomly drawn for each trial from a Normal distribution with mean 0 and standard deviation of .20. Starting values of the weights were 0.15. The simulation was run separately for each condition, and in each condition each trial was repeated twice. After each condition, the “Measures” section was run.

During learning, I used learning rate = 0.25 with linear activation updating and number of internal cycles = 1. The external activation levels listed in Table 2 received additional random noise generated from a Normal distribution with mean 0 and standard deviation 0.20. The starting values of the weights were set at 0.15. This reflects the idea that perceivers typically start with the assumption that the actor possesses the inferred traits and motives to some minor degree.

At the end of each simulated experimental condition, to simulate the empirical dependent measures, test trials were run by prompting certain nodes of interest (i.e., turning on their activation), and the resulting output activation in other nodes was recorded. For instance, to test trait inferences, the actor node was turned on and the resulting activation of the trait-implying behavior node (without any additional external activation) was read off. Similar test procedures for the other dependent variables can be seen from the bottom panel of Table 2.

The simulations were repeated for 100 “participants”, and the results were then averaged. These simulations were then verified by comparing the resulting mean test activations with observed experimental data. Given that the resulting activation values and experimental results are difficult to compare quantitatively, I examined only the general pattern of activations and projected them visually onto the observed data (i.e., I re-scaled the obtained mean test activations by linear regression with a positive slope).

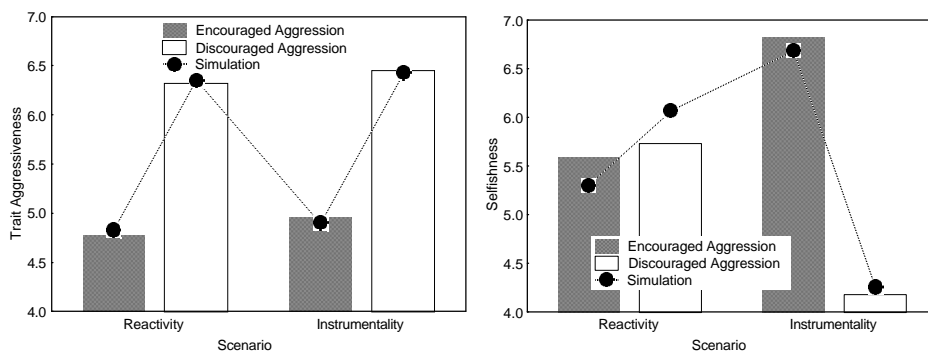
5.2. Results

As can be seen from Figure 2, the simulated values reproduced the observed ratings very closely. The aggressiveness traits failed to show a significant interaction between type of scenario, and revealed only the hypothesized main effect of situation, $F(1, 396) = 465.34, p < 0.001$. This reflects discounting of trait aggressiveness regardless of the scenario (see Figure 2, first panel). In contrast, the ratings of selfishness, motive to revenge and motive to gain money, showed a significant interaction, $F_s(1, 396) = 100.37\text{—}1214.99, p_s < 0.001$. As can be seen in Figure 2 (second panel), consistent with the empirical data, although simulated selfishness was discounted in the revenge condition, it was not discounted in the instrumental aggression condition. For motives of revenge and gain (Figure 2, bottom panels), the simulation generally showed

the same pattern of differences between conditions that encouraged and discouraged aggression like in the replication study.

6. General Conclusion

The present simulation was able to resolve an apparent paradox between trait and moral inferences found in my replication of Reeder et al (2002). By extending earlier connectionist architectures of person perception (Smith & DeCoster, 1998; Van Overwalle & Labiouse, 2004) with the motives people may infer about the actor's behavior, discounting was consistently applied for trait ratings given various situational forces, but was absent for moral (i.e., selfishness) ratings given some types of situational forces. In other words, the traits ratings did not reveal the fundamental attribution bias, whereas the moral ratings did. A crucial feature that made it possible to simulate the differences between trait and moral inferences was that the selfishness and motive measures included either the motive to revenge or the motive to gain. This resulted in an asymmetric discounting pattern in the reactivity and instrumentality scenarios. In contrast, the trait measure did not involve these motives (although traits were affected by these motives during the learning history). Consequently, this resulted in a more symmetric discounting pattern across scenarios.



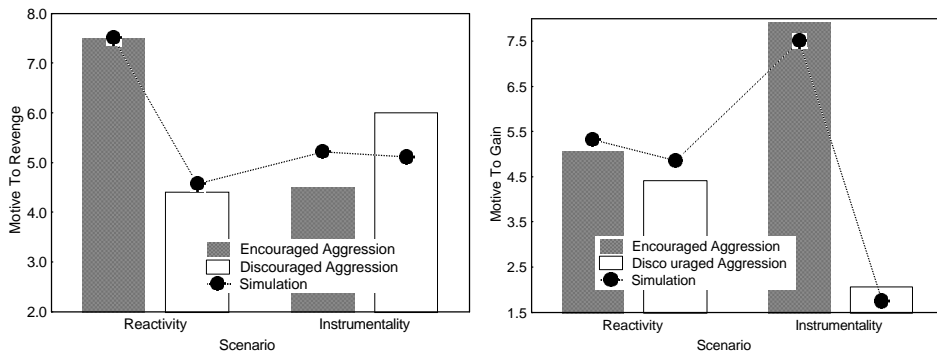


Figure 2. Perception of an Aggressive Actor as a Function of Scenario and Situation. Observed and Simulated Values from the Replication Study.

References

1. Gilbert, D. T. (1989) Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.) *Unintended thought*. New York, NY: Guilford.
2. Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology*, 30, 407–455.
3. Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impression as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review*, 107, 914-942.
4. Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.
5. McClelland, J. L. & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology*, 114, 159–188.
6. Read, S. J., & Montoya, J. A. (1999). An autoassociative model of causal reasoning and causal learning: Reply to Van Overwalle's critique of Read and Marcus-Newhall (1993). *Journal of Personality and Social Psychology*, 76, 728–742.
7. Reeder, G. D., Kumar, S., Hesson-McInnis, M. S. and Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, 83, 789–803.

8. Reeder, G. D., Vonk, R., Ronk, M. J., & Ham, J. (2003). *Dispositional attribution: Multiple inferences about motive-related traits*. Unpublished Manuscript.
9. Shultz, T. & Lepper, M. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review*, 2, 219-240.
10. Smith, E. R. & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology*, 74, 21—35.
11. Van Overwalle, F. (1998) Causal Explanation as Constraint Satisfaction: A Critique and a Feedforward Connectionist Alternative. *Journal of Personality and Social Psychology*, 74, 312-328.
12. Van Overwalle, F., & Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review*, 3, 204—231.
13. Van Overwalle, F., & Labiouse, C. (2004) A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review*, 8, 28—61.
14. Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C. & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review*, 110, 536-563.