Reasoning about social choices and social relationships

Alan Jern
jern@rose-hulman.edu
Department of Humanities and Social Sciences
Rose-Hulman Institute of Technology

Charles Kemp
ckemp@cmu.edu
Department of Psychology
Carnegie Mellon University

Abstract

We study inferences about social choices—choices that affect people besides the chooser. Social choices depend on the relationships between the people involved: for example, whether they are friends, strangers, or enemies. We propose that these different social relationships correspond to different ways in which the chooser weights another person's utility relative to her own. We describe a probabilistic model of social reasoning that incorporates this notion of weighted utility, and evaluate it in an experiment in which participants made inferences about others' social choices. The results support our probabilistic model and expose some of the assumptions that people tend to make when reasoning about social choices.

Keywords: social cognition; social reasoning; folk social psychology; probabilistic models

People frequently engage in folk psychological reasoning. We explain and predict other people's behavior and draw inferences about other people's thoughts and feelings. This reasoning sometimes depends on knowledge of social relationships. For example, if you know that Alice and Bob are friends, you might predict that Alice would be willing to make a financial sacrifice to help Bob, perhaps by offering him a loan. If you know that Alice and Bob are enemies, you might predict that Alice would not be willing to make a financial sacrifice to benefit Bob, but she might be willing to make a financial sacrifice to harm Bob, perhaps by turning down a mutually beneficial business opportunity.

Alice's choices above are instances of *social choices*—choices that affect people besides the chooser. Her choices each result in a cost to herself, but also a benefit or cost to Bob. These examples illustrate how knowledge of the relationship between two people can inform expectations about the social choices that they will make. Conversely, observing a social choice may allow us to infer something about the relationship between the people involved. Despite the fact that people commonly reason about social choices and social relationships, there are few formal proposals about how people perform this sort of reasoning (Haslam, 1994). We suggest in this paper that inferences about social choices and relationships can be viewed as a kind of probabilistic reasoning.

Previous research has explored how people reason about other people's non-social choices, like choosing which shirt to buy. Standard choice models can be used to predict the choices that follow from a given set of preferences, and "inverting" these models provides a way to reason backward and infer the preferences that likely motivated an observed choice. Several studies have shown that this inverse reasoning approach accounts well for experiments that focus on reasoning about non-social choices (Lucas et al., 2014; Bergen, Evans, & Tenenbaum, 2010; Jern & Kemp, 2011; Jern, Lucas, & Kemp, 2011). In social settings, however, the chooser's utility may depend on the utility experienced by others. One way to capture this dependence is to suppose that the chooser's utility function is a weighted combination of the utilities directly experienced by all affected individuals (Wyer, 1969; McClintock, 1972; Griesinger & Livingston, Jr., 1973). We propose that people represent different social relationships as different utility weighting functions. We show that combining this proposal with the inverse reasoning approach can account for a wide array of social inferences, including inferences about whether a pair of people are more likely to be friends, enemies, or strangers.

Our proposal is conceptually related to previous computational approaches that have been used to explain how people infer social goals like "helping" (Baker, Goodman, & Tenenbaum, 2008; Ullman et al., 2009). These approaches have focused on how people reason about sequences of actions that extend through time and space. By contrast, we explore one of the simplest possible settings that supports inferences about social choices and how such choices are affected by social relationships.

The next two sections introduce our formal approach in more detail. We then evaluate our approach in an experiment in which participants made several kinds of inferences about social choices.

A social choice model

We propose that people reason about social choices by inverting a simple model of how utilities give rise to social choices. Consistent with previous approaches (Train, 2009), we assume that utilities are additive and that people tend to choose options with greater utilities. With social choices, the notion of utility can be confusing because utility is not necessarily identical to a direct reward. For example, if Alice's choice can either benefit or harm Bob, Alice's utility may depend on the effect her choice has on Bob in addition to any benefit her choice provides for herself. To alleviate this confusion, we will refer to the utilities assigned to rewards or payouts as *direct utilities* and will use the term "utility" to refer to a chooser's total utility.

How much utility a chooser assigns to different options in a social choice depends on how the chooser weights the direct utilities of everyone affected by the choice. Here we assume that there is only one other person affected by the choice, but our approach can be straightforwardly extended to include any number of people. We will henceforth refer to the chooser as Alice and the person affected by the choice as Bob. Let w_A be the weight that Alice assigns to her own direct utility and let w_B be the weight that Alice assigns to Bob's direct utility. We constrain these weights to sum to 1: $w_A + w_B = 1$. When $w_B = 0$, Alice does not take Bob's direct utility into account at all. To capture the idea that Alice may either want to help or harm Bob, we include another parameter $\gamma_B \in \{-1, +1\}$ that specifies the polarity that Alice assigns to Bob's utility. When $\gamma_B = -1$, Alice's utility increases as Bob's direct utility decreases, and when $\gamma_B = 1$, Alice's utility increases as Bob's direct utility increases. The former case might apply when Alice and Bob are enemies and the latter case might apply when they are friends.

We can now specify how Alice's utility is computed. We will use U_A to represent her total utility and u_A and u_B to represent direct utilities for Alice and Bob. Consistent with previous research (Wyer, 1969; McClintock, 1972; Griesinger & Livingston, Jr., 1973), we define Alice's total utility as a linear combination of direct utilities:

$$U_A = w_A \cdot u_A + \gamma_B \cdot w_B \cdot u_B.$$

In some cases, Alice will not know Bob's direct utility before making a choice. In these cases, Alice's (possibly false) beliefs about u_B will influence her choices. We therefore assume that u_B represents Alice's belief about Bob's direct utility, rather than the true value of Bob's direct utility.

We will refer to a particular setting of these parameters as Alice's social attitude toward Bob. Figure 1a shows all the possible settings of these parameters. The cyan line on the top half corresponds to cases in which $\gamma_B = 1$ and the green line on the bottom half corresponds to cases in which $\gamma_B = -1$. Some of these social attitudes are labeled in the figure using English adjectives. For example, point $(0,1)$ is labeled "altruistic" because in this case Alice completely discounts her own direct utility in favor of Bob's. Figure 1b shows a set of options along the blue arc that might be available to Alice. Each option results in a payout to both Alice and Bob. Every point along the lines in Figure 1a corresponds uniquely to a preferred option along the arc in Figure 1b. Some of these options are identified by the social attitudes that would lead Alice to prefer them. For example, if Alice is altruistic, she will choose the $(50,100)$ option, which provides the greatest payout to Bob.

To complete the social choice model, we must specify how Alice selects an option. We will refer to a function that specifies how choices are made as a *choice function*. When modeling the experiment discussed later, we consider two common choice functions. The first is a utility-maximizing choice function, which specifies that Alice will select the option that maximizes U_A. We will call the second a utility-matching function because it specifies that Alice will select options probabilistically in proportion to their utilities.

Representing social relationships

We now show how it is possible to define different social relationships in terms of the social attitude parameters defined in the previous section. In this paper, we focus on three relationships: friends, enemies, and strangers. The left side of

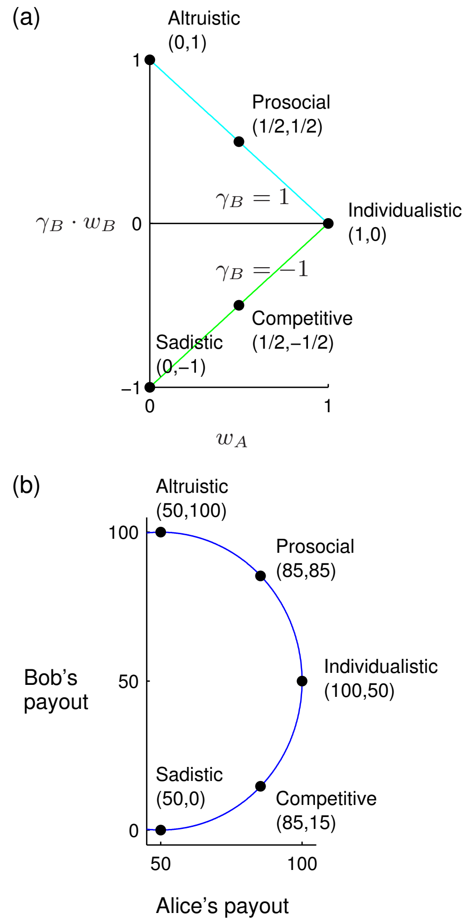

Figure 1: Social attitudes and social choices. (a) The two line segments show all possible settings of parameters w_A, w_B, and γ_B. Some settings correspond to the social attitudes indicated by the labels. (b) A set of options available to Alice. Alice's preferred option depends on her social attitude toward Bob. This figure is adapted from Liebrand (1984) and Murphy and Ackermann (2012).

Table 1 ("Basic relationship model") summarizes the commonsense assumptions we make about these relationships. If Alice and Bob are friends, we assume that Alice assigns a positive polarity to Bob's direct utility (i.e, $\gamma_B = 1$). If Alice and Bob are enemies, we assume that Alice assigns a negative polarity to Bob's direct utility (i.e., $\gamma_B = -1$). Note that for both friends and enemies, we make no assumption about the relative weights Alice assigns to her own and Bob's direct utilities. If Alice and Bob are strangers, we assume that Alice weights her own direct utility more than Bob's, but we make no assumption about the polarity Alice assigns to Bob's direct utility.

We now show how performing probabilistic inference using the social choice model in the previous section can be used to account for inferences about social relationships. Suppose you observe Alice make a social choice. For example, in the experiment described in the next section, we

	Basic relationship model		No relationship model		Augmented relationship model	
Relationship	w_A and w_B	γ_B	w_A and w_B	γ_B	w_A and w_B	γ_B
Friends	–	1	–	–	$w_A < w_B$	1
Strangers	$w_A > w_B$	–	–	–	$w_A > w_B$	1
Enemies	–	-1	–	–	$w_A > w_B$	-1

Table 1: Three social relationships defined in terms of the social attitude parameters w_A, w_B, and γ_B. This table shows assumptions made by three models described in the text. The "–" entries indicate no assumptions.

consider a simple social choice in which Alice chooses a type of candy that will be given to both her and Bob. Let Alice's choice be denoted by c_A. Using the assumptions just described, we can apply Bayes' rule to compute probabilities of each relationship after observing c_A:

$$\begin{aligned} P(\text{Friends}|c_A) &\propto P(c_A|\text{Friends})P(\text{Friends}) \\ &= P(c_A|\gamma_B = 1)P(\text{Friends}) \qquad (1) \\ P(\text{Strangers}|c_A) &\propto P(c_A|\text{Strangers})P(\text{Strangers}) \\ &= P(c_A|w_A > w_B)P(\text{Strangers}) \qquad (2) \\ P(\text{Enemies}|c_A) &\propto P(c_A|\text{Enemies})P(\text{Enemies}) \\ &= P(c_A|\gamma_B = -1)P(\text{Enemies}) \qquad (3) \end{aligned}$$

Computing the likelihood terms in these equations, like $P(c_A|\gamma_B = 1)$ in Equation 1, requires integrating over the values of the other parameters. For example, $P(c_A|\gamma_B = 1) = \int_{w_A}\int_{w_B} P(c_A|w_A, w_B, \gamma_B = 1)P(w_A, w_B)\mathrm{d}w_B\mathrm{d}w_A$. We assume a uniform prior distribution over the parameters w_A, w_B, and γ_B, and compute $P(c_A|w_A, w_B, \gamma_B)$ by applying the social choice model.

Experiment

We evaluated our formal approach by comparing its predictions to people's inferences in an experiment. Our experiment involved a simple social choice in which Alice chooses between two types of candy and the candy she chooses is given to both her and Bob. Although this choice is simple, it depends on three factors: Alice's candy preferences, Alice's beliefs about Bob's candy preferences, and Alice's social attitude toward Bob. In other words, this simple choice captures one important feature of real-life social choices: Alice must balance her own interests against Bob's interests.

We provided participants with three of the following four pieces of information and asked them to infer the fourth: (1) the type of candy Alice likes best, (2) the type of candy Alice believes that Bob likes best, (3) Alice's and Bob's relationship, and (4) Alice's choice of candy—either Candy 1 or Candy 2. Alice and Bob can each prefer either Candy 1 or Candy 2, Alice can choose Candy 1 or Candy 2, and we considered three possible relationships between Alice and Bob: friends, strangers, and enemies. Eliminating one of the four pieces of information and enumerating all possible values of the remaining pieces of information results in 22 distinct inference problems, each of which constituted an experimental condition. These conditions are shown in Figure 2a, where the labels above the plots indicate what information was provided in that condition. In six conditions (Figure 2a.i), participants inferred Alice's preference; in six conditions (2a.ii), participants inferred Alice's belief about Bob's preference; in six conditions (2a.iii), participants predicted Alice's social choice; and in four conditions (2a.iv), participants inferred Alice's and Bob's relationship.

Model

We considered three models of this task. All models assume that Alice is more likely to choose options with greater utility and that the utility she assigns to each option is a weighted function of her and Bob's direct utilities. The models differ with respect to the assumptions they make about how people think about friends, strangers, and enemies. Each model's assumptions are shown in Table 1.

The first model, which we will refer to as the *basic relationship model*, makes three assumptions that were previously described and are shown in the left column of Table 1. We call this model the basic relationship model because the few assumptions that it makes seem obligatory in order to capture commonsense expectations about friends, strangers, and enemies.

In order to evaluate whether the assumptions made by the basic relationship model are needed to account for people's inferences, we considered a baseline model that places uniform prior distributions over each model parameter. Although this *no relationship* model makes no assumptions at all about the differences between the three relationships, it still assumes that Alice is more likely to choose options with greater utility. As a result, there are some situations in which the model should make sensible inferences about Alice.

The third model in Table 1 adds three assumptions to the basic relationship model. We will describe this third model after presenting our data.

All models assume a uniform prior distribution on the three relationships, meaning that all three types of relationships are initially judged to be equally probable. This assumption is unlikely to be true in general, but the cover story of our experiment suggested that, in the context of the experiment, the three types of relationships were equally common. We assumed that Alice assigns a generally positive utility to each type of candy. To be consistent with previous research (Lucas et al., 2014; Jern et al., 2011), both models also assume normal prior distributions over utilities ($u_i \sim \mathcal{N}(2,4)$), but the models make nearly identical predictions given other distributional assumptions (e.g., $u_i \sim \text{Uniform}(0,10)$). We generated one set of predictions for each model using a utility-maximizing choice function and one set of predictions using a utility-matching choice function.

Method

Participants 80 participants were recruited from the Amazon Mechanical Turk website. They were paid for their participation.

Design We used a mixed design in which each participant completed all of the conditions for one type of inference. That is, each participant completed all of the conditions in one row of Figure 2a. Participants were randomly assigned to inference type and completed the conditions in a random order.

Procedure The experiment was completed online. Each condition appeared on a separate page. Participants were first told that a group of researchers had conducted a study to see how people in different relationships make choices that affect others. They were told that the study involved pairs of people and that some of these pairs of people were friends, some were strangers, and some were enemies. Finally, they were told that the choices in the study were about two different types of candy, Candy 1 and Candy 2, and that all of the people in the experiment had been allowed to try both types of candy before making any choices.

In each condition, participants were presented with three of the following four pieces of information about one pair of people in the fictional study.

- Alice and Bob are [friends/strangers/enemies].
- Alice was asked which type of candy she liked best. Alice liked [Candy 1/Candy 2] best.
- Alice was asked which type of candy she thought Bob liked best. Alice thought that Bob liked [Candy 1/Candy 2] best.
- Alice was asked to pick a candy that would be given to both Alice and Bob. Alice picked [Candy 1/Candy 2].

The information in brackets depended on the condition. Whether Alice liked Candy 1 or Candy 2 best was randomized across participants. The actual names used in the experiment were also randomized.

In the inference phase of each condition, participants responded to one of the following four sets of questions, depending on the inference condition.

- **Alice's preference**: How likely is it that ... (1) Alice liked Candy 1 best? (2) Alice liked Candy 2 best?
- **Alice's belief**: How likely is it that ... (1) Alice thinks Bob likes Candy 1 best? (2) Alice thinks Bob likes Candy 2 best?
- **Social choice**: Alice was asked to pick a candy that would be given to both Alice and Bob. How likely is it that ... (1) Alice picked Candy 1 for both Alice and Bob? (2) Alice picked Candy 2 for both Alice and Bob?
- **Relationship**: How likely is it that Alice and Bob are ... (1) Friends? (2) Strangers? (3) Enemies?

Participants answered each question using a slider that spanned from 0 (very unlikely) to 100 (very likely).

Results

Model predictions The information provided about Alice was encoded in the models as follows. Alice's preference was captured by placing a constraint on her direct utilities. For example, if Alice preferred Candy 1, the models assumed that $u_A^1 > u_A^2$, where the subscript indicates the person and the superscript indicates the candy. Similarly, if Alice believed that Bob preferred Candy 1, the models assumed that $u_B^1 > u_B^2$.

The model predictions were computed by conditioning on the provided information. For example, in one condition, the provided information stated that Alice prefers Candy 1, Alice believes that Bob prefers Candy 2, and Alice chose Candy 2. The probability that Alice and Bob are friends can therefore be computed as follows:

$$P(\text{Friends}|u_A^1 > u_A^2, u_B^1 < u_B^2, c_A = 2) \propto \\ P(c_A = 2|\text{Friends}, u_A^1 > u_A^2, u_B^1 < u_B^2)P(\text{Friends}).$$

The first term on the right-hand side of the equation can be expanded by applying the model's definition of the friends relationship and then applying the appropriate choice function. Predictions in the preference inference conditions were similarly generated by computing $P(u_A^1 > u_A^2|\text{Relationship}, \text{Bob's preference}, c_A)$ for each condition. Predictions in the belief inference conditions were generated by computing $P(u_B^1 > u_B^2|\text{Relationship}, \text{Alice's preference}, c_A)$, and predictions in the choice prediction conditions were generated by computing $P(c_A|\text{Relationship}, \text{Alice's preference}, \text{Bob's preference})$. We generated 200,000 samples from the appropriate distribution in each condition by importance sampling with samples drawn from the corresponding prior distributions.

Human judgments Participants' ratings for each condition were normalized so that their ratings summed to 1. Figure 2a compares ratings for all conditions with the predictions of the utility-matching basic relationship model. These plots suggest that the commonsense assumptions made by the basic relationship model are sufficient to account for people's inferences in most conditions.

The overall performance of the basic relationship and no relationship models is shown in Figures 2b.i and 2b.ii. Figure 2b.i suggests that the basic relationship model predicts people's quantitative judgments quite well. The model performs significantly better than the no relationship model ($z = 3.04$, $p < .01$), suggesting that participants did distinguish between friend, stranger, and enemy relationships, and expected these relationships to influence people's social choices. The utility-maximizing versions of both models performed worse ($r = 0.88$ and $r = 0.72$) than the utility-matching basic rela-

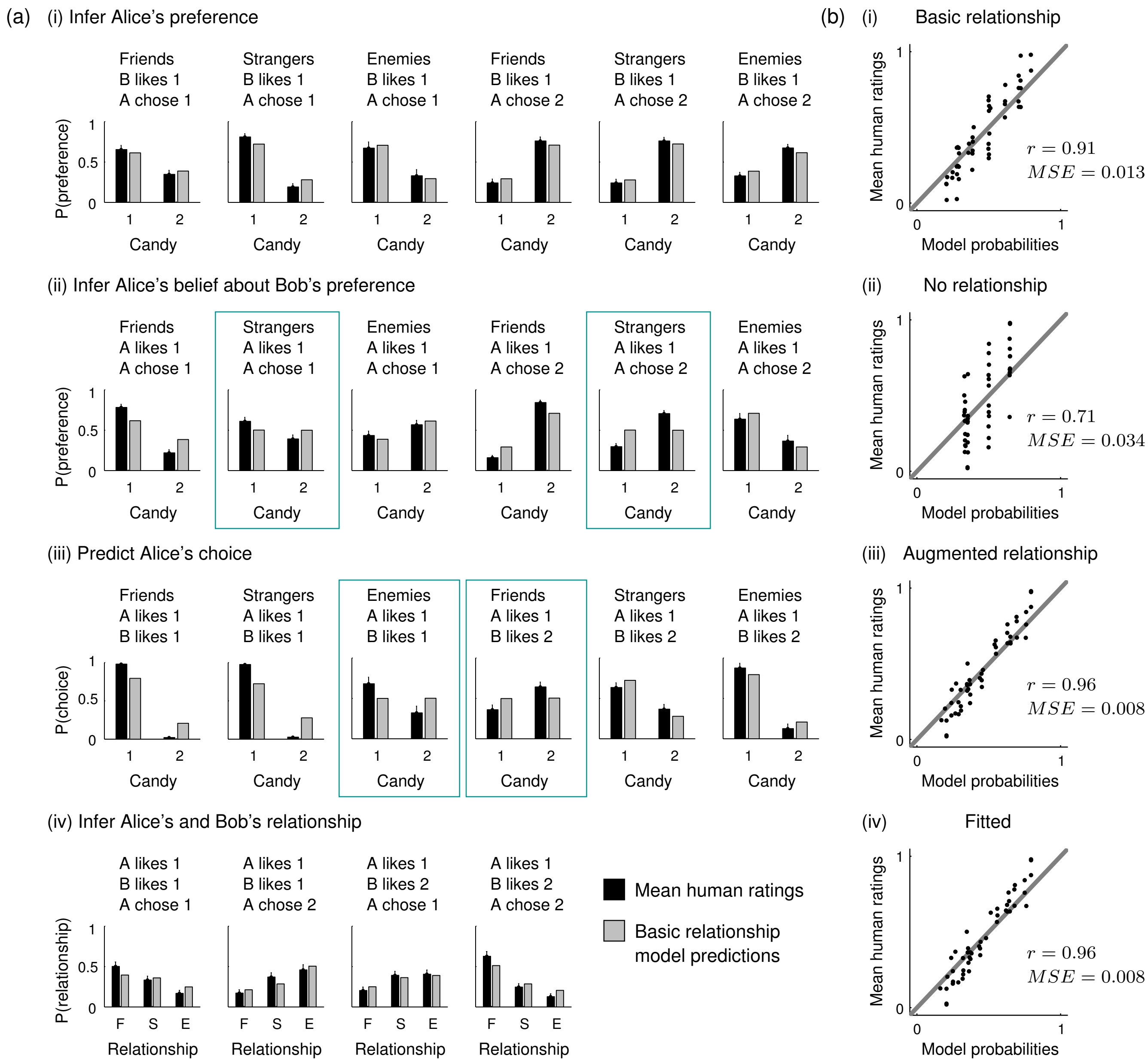

Figure 2: Experiment results and model predictions. (a) Comparison between the utility-matching basic relationship model predictions and mean human ratings. Each row shows results for one type of inference. The labels at the top of each plot indicate the information that the model and participants were provided with. In the last row, F = friends, S = strangers, and E = enemies. The error bars for the mean human ratings indicate standard errors of the mean. The conditions in boxes motivated the augmented relationship model discussed in the text. (b) Overall performance of the four models discussed in the text. Predictions are for utility-matching choice functions in all cases. The gray lines in the plots indicate perfect correspondence between the model predictions and human ratings. MSE = mean squared error.

tionship model. Therefore, we will not discuss utility maximization any further.

Despite the overall accuracy of the basic relationship model, people's inferences in several conditions, indicated by boxes in Figure 2a, suggest that people made some additional assumptions that are not captured by the model. For example, consider the condition in Figure 2a.iii in which Alice and Bob are enemies, Alice likes Candy 1 best, and Alice believes that Bob likes Candy 1 best. The model predicts that it is equally probable that Alice will choose Candy 1 and Candy 2 because the model does not make any assumptions about whose direct utility Alice will weight more heavily. Alice might weight her own direct utility more heavily and choose Candy 1, or she might weight Bob's direct utility more heavily and choose Candy 2 because the pair are enemies. By contrast, participants judged it more likely that Alice would choose Candy

1, consistent with an expectation that Alice tends to weight her own direct utility more heavily. The other highlighted conditions in Figure 2a suggest that people expected Alice to weight Bob's direct utility more than her own when they were friends, and expected Alice to assign a positive polarity to Bob's direct utility when they were strangers. These expectations are broadly consistent with how people in different relationships actually make social choices (Loewenstein, Thompson, & Bazerman, 1989). We therefore considered an additional model that captures these expectations.

The augmented relationship model The third column of Table 1 adds the three assumptions just described to the three assumptions of the basic relationship model. We refer to the resulting model as the *augmented relationship model*.

The overall performance of the augmented relationship model (with a utility-matching choice function) is shown in Figure 2b.iii. The model predicts participants' judgments ($r = 0.96$) significantly better than the utility-matching basic relationship model ($z = 1.98, p < .05$). Although not shown, the model also captures the qualitative effects in the highlighted conditions in Figure 2a that the basic relationship model was unable to capture.

To obtain an upper bound on model performance, we also considered a model in which all of the social attitude parameters were fit to the data. For each relationship, we fit the parameters w_A, w_B, and γ_B. We generated model predictions for every combination of parameters, with direct utility weights varying in increments of 0.1, and fit the predictions to the complete set of mean participant ratings. We chose the values of the parameters that produced the largest correlation coefficient between model predictions and mean participant ratings. The overall performance of the fitted relationship model is shown in Figure 2b.iv. As the figure shows, the fitted model offers virtually no performance gains over the augmented relationship model, suggesting that the relationship assumptions made by the augmented relationship model are sufficient to capture people's inferences in our task.

Conclusion

Our data support the idea that different social relationships correspond to different ways of weighting utility functions. Three commonsense assumptions about the nature of these weights for different relationships allowed us to predict peoples social inferences in a large number of cases fairly accurately. Our experimental results also revealed three additional assumptions that people seem to make about how friends, strangers, and enemies will behave toward one another.

This paper focused on one of the simplest examples of folk social psychology: reasoning about a choice that affects one person in addition to the chooser. Future research should explore how people reason about more complex social situations. For example, situations in which two people simultaneously make social choices that affect one another introduce a recursive aspect to the social decision-making process that is not captured by the present version of our model. Additionally, social choices are often influenced by social norms and other factors that go beyond the direct costs and benefits to everyone affected. An important goal for research in both cognitive science and social cognition is to develop a general computational account of folk social psychology that applies to situations like these. Our work suggests that utility weighting functions and probabilistic inference are two principles that can contribute to such an account.

Acknowledgments This work was supported by the Pittsburgh Life Sciences Greenhouse Opportunity Fund, NSF Grant CDI-0835797, and NIMH Training Grant T32MH019983.

References

Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based social goal inference. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Bergen, L., Evans, O. R., & Tenenbaum, J. B. (2010). Learning structured preferences. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

Griesinger, D. W., & Livingston, Jr., J. W. (1973). Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, *18*(3), 173–188.

Haslam, N. (1994). Mental representation of social relationships: Dimensions, laws, or categories? *Journal of Personality and Social Psychology*, *67*(4), 575–584.

Jern, A., & Kemp, C. (2011). Decision factors that support preference learning. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

Jern, A., Lucas, C. G., & Kemp, C. (2011). Evaluating the inverse decision-making approach to preference learning. In *Advances in Neural Information Processing Systems 24.*

Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behavior in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*(3), 239–264.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, *57*(3), 426–441.

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., et al. (2014). Child as econometrician: A rational model of preference understanding in children. *PLOS ONE*, *9*(3), e92160.

McClintock, C. G. (1972). Social motivation—a set of propositions. *Behavioral Science*, *17*(5), 438–454.

Murphy, R. O., & Ackermann, K. A. (2012). *A review of measurement methods for social preferences.* (Working paper)

Train, K. (2009). *Discrete choice models with simulation* (2nd ed.). New York, NY: Cambridge University Press.

Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in Neural Information Processing Systems 22.*

Wyer, R. S. (1969). Prediction of behavior in two-person games. *Journal of Personality and Social Psychology*, *13*(3), 222–238.