Improving Confidence in the Estimation of Values and Norms

Cavalcante Siebert, L.; Mercuur, R.A.; Dignum, M.V.; van den Hoven, M.J.; Jonker, C.M.

DOI
10.1007/978-3-030-72376-7_6

Publication date
2020

Document Version
Final published version

Published in
Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIII - International Workshops COIN 2017 and COINE 2020, Revised Selected Papers

Citation (APA)
Cavalcante Siebert, L., Mercuur, R. A., Dignum, M. V., van den Hoven, M. J., & Jonker, C. M. (2020). Improving Confidence in the Estimation of Values and Norms. In A. Aler Tubella, S. Cranefield, C. Frantz, F. Meneguzzi, & W. Vasconcelos (Eds.), *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIII - International Workshops COIN 2017 and COINE 2020, Revised Selected Papers: International Workshops COIN 2017 and COINE 2020 Sao Paulo, Brazil, May 8–9, 2017 and Virtual Event, May 9, 2020 Revised Selected Papers* (pp. 98-113). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12298 LNAI). ArXiv. https://doi.org/10.1007/978-3-030-72376-7_6

Improving Confidence in the Estimation of Values and Norms

Luciano Cavalcante Siebert[1,2(✉)], Rijk Mercuur[3], Virginia Dignum[3,4], Jeroen van den Hoven[2,3], and Catholijn Jonker[1,2,5]

[1] Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, The Netherlands
l.cavalcantesiebert@tudelft.nl
[2] AiTech, Delft University of Technology, Delft, The Netherlands
[3] Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands
[4] Department of Computing Sciences, Umeå University, Umeå, Sweden
[5] Leiden Institute of Advance Computer Science, Leiden University, Leiden, The Netherlands

Abstract. Autonomous agents (AA) will increasingly be interacting with us in our daily lives. While we want the benefits attached to AAs, it is essential that their behavior is aligned with our values and norms. Hence, an AA will need to estimate the values and norms of the humans it interacts with, which is not a straightforward task when solely observing an agent's behavior. This paper analyses to what extent an AA is able to estimate the values and norms of a simulated human agent (SHA) based on its actions in the ultimatum game. We present two methods to reduce ambiguity in profiling the SHAs: one based on search space exploration and another based on counterfactual analysis. We found that both methods are able to increase the confidence in estimating human values and norms, but differ in their applicability, the latter being more efficient when the number of interactions with the agent is to be minimized. These insights are useful to improve the alignment of AAs with human values and norms.

Keywords: Autonomous agents · Values · Norms · Ultimatum game

1 Introduction

As autonomous agents (AAs) become more pervasive in our daily lives, there is a growing need to reduce the risk of undesired impacts on our society [2,12]. Hence, we need design and engineering approaches that consider the implications of ethically relevant decision-making by machines, understanding the AA as part of a socio-technical system [8, p.48]. For this, we need to ensure that the decisions and actions made by the AA are aligned with the stakeholders' values ("what one finds important in life" [23]) and norms (what is standard, acceptable or permissible behavior in a group or society [10]).

A. Aler Tubella et al. (Eds.): COIN 2017/COINE 2020, LNAI 12298, pp. 98–113, 2021.
https://doi.org/10.1007/978-3-030-72376-7_6

Values and norms can be seen as criteria for decision-making: values relating to more abstract and context-independent ideals, and norms as more concrete context-dependent rules. Furthermore, since humans use values and norms in their explanations, the use of these concepts allows for a better explanation of the AA [15–17]. People have a common base system of values and they are relatively stable over the life span of a person [4], however, values and norms vary significantly for each person and socio-cultural environment, making it difficult or even impossible to write down precise rules describing them. One approach to deal with this variance is to treat aligning the actions of the AA with human values as a learning problem [13,27]. However, estimating values and norms solely based on observed behavior may lead to ambiguous results. Different relative preferences towards values and norms can bring about the same behavior, i.e. there are many "reasons why" that might motivate a given observed behavior (ambiguity problem). In case that an agent's actions are driven by a "wrong" set of values and norms, strong ethical consequences may befall.

This paper studies the conditions under which AAs are able to make confident estimates of one's preference on values and norms from observed behavior. We studied the behavior of simulated agents driven by both values and social norms in the Ultimatum Game (UG), specifically in the proposer role. [20] has used the UG to study the relative importance of values across cultures. We focus on the relative importance of values and norms that guide individual decision making, aiming to provide useful insights to improve the alignment of AAs with human values and norms.

The main contribution of this paper is twofold. First, we propose an agent-based model to play the UG, expanded from [16], which uses both values and norms for determining the actions of the agents. Second, we present a method for estimating an agent's relative preferences to a given set of values and, if necessary, to improve the confidence in these estimations. This improvement is realized by interacting with the proposer agent in the UG, either by a free exploration of the search space or by the use of counterfactuals.

The remainder of the paper is structured as follows. The following section presents a brief context about the UG and how we model values and norms in this context. Section 3 presents both the method to estimate the relative preference attributed to an agent's values and norm, and the methods to reduce ambiguity on the estimation. Section 4 presents the main results of this research and Sect. 5 discusses these results. Finally, Sect. 6 presents the conclusions.

2 Values and Norms in the Ultimatum Game

2.1 Scenario: The Ultimatum Game

We simulate the behavior of the agents in the UG, first introduced by [11]. In the UG, two players negotiate over a fixed amount of money (the 'pie'). The proposer demands a portion of the pie, while the remainder is offered to the responder. The responder chooses to accept or reject this proposed split. If the responder chooses to 'accept', the proposed split is implemented. If the responder chooses

to 'reject', both players get no money. The UG thus provides an environment where players have to make decisions about money based on their own judgment.

The remainder of the paper uses a version of the UG with the following specifics:

- An experiment has 32 players: 16 proposers and 16 responders.
- One experiment has 20 rounds. One round in the UG comprises one demand for each proposer and one reply for each matched responder.
- The players are paired to a different player each round but do not change roles.
- Players are anonymous to each other.
- The pie size P is 1000[1].

These specifics are chosen for pragmatic reasons: the paper uses an empirical dataset based on these specifics [3] and builds upon a model based on these specifics [16].

This paper models the decision-making of humans in the UG by using values and norms, but there have been many different approaches to it since its first appearance in [11]. One reason for its popularity is that the UG is a classical example of where the canonical game-theoretical agent (i.e., the *homo economics* that only cares about maximizing its direct own welfare) falls short. The *homo economicus* only cares about its direct welfare and thus will accept any positive offer in the UG. Humans, in contrast, reject offers as high as 40% of the pie [21]. Behavioral economics has aimed to explain humans by incorporating learning [25], reputation [9] or other-regarding preferences. Recently, [16] used values and norms to explain UG behavior. Using a series of agent-based simulation experiments, [16] showed that the model produces aggregate behavior that falls within the 95% confidence interval wherein human behavior lies more often than the other tested agent models (e.g., a learning *homo economicus* model). Moreover, the model uses concepts that humans use in their explanations. Thus because the model has been shown to reproduce human behavior and uses explainable concepts, the model of [16] provides a starting point to estimate preference towards human values and norms for value alignment.

2.2 Simulated Human Agent (SHA)

The simulated human agent (SHA) represents the human in the UG whose values and norms we aim to estimate. The model for the SHA is based on the value-based model and norm-based model presented in [16]. These models focus on the aggregate properties that emerge from pairing these agents in the UG. However, this paper focuses on estimating relative preference towards values and norms of individual agents. Therefore, we are interested in an agent model where one agent uses *both* values and norms. The remainder of this section presents the SHA model in three parts: the value-based agent (from [16]), the norm-based agent (extension of [16]) and an agent that uses both values and norms (new).

[1] For ease of presentation, we chose to present P with no monetary unit. Empirical work [21] shows that the effect of the pie size is relatively small.

Value-Based Agent. The value-based agent uses the importance an agent attributes to its values to determine what an agent (based on its values) should demand. [16] focuses on two values that are relevant in the UG: wealth and fairness and define di_a as the difference in importance an agent attributes to wealth versus fairness. Based on research by [26] and data on the UG [3], they then state a number of requirements for a function (see [16] for more details). First, they assume a (perfect) negative correlation between wealth and fairness. Second, valuing wealth is defined as wanting more money. Third, valuing fairness is defined as wanting an equal split of the pie. Fourth, agents should differ in how much money they demand, but not demand below 0.5P. The functions that map di_a to the demand $d \in \{0...P\} \subset \mathbb{Z}$ are given by

$$valueDemand(di_a) = \underset{d \in \{0...P\} \subset \mathbb{Z}}{\arg\max}\ u(d, di_a, P) \tag{1}$$

and

$$u(d, di_a, P) = -\frac{1.0 + 0.5di_a}{\frac{d}{P} + 0.5} - \frac{1.0 - 0.5di_a}{\frac{|0.5P - d|}{0.5P} + 0.5} \tag{2}$$

The agent thus calculates the utility for each demand d and returns the demand with the maximum utility. [16] shows these functions fulfill the stated requirements. By using (1) and (2), we enable our agent to choose a demand following theories on values [26] and [3] empirical results on the UG.

Norm-Based Agent. The norm-based agent uses the replies it observes from the other agent to determine what an agent (based on its norms) should demand. In [16], following theories by [5] and [10], it is stated that a norm exists for a particular person when that person perceives what most other people do or expect it. [16] provides a translation from this definition to a model for the UG. In the UG, the proposer cannot see what other proposers do, because the proposer is only paired with responders. Therefore, the proposer uses the responses of the responder to form an idea of what is expected (i.e., the norm). [16] provides the following function to map the observed responses (OR_{ak}) seen by agent a up to the current round k to a demand $d \in \{0...P\} \subset \mathbb{Z}$,

$$normDemand(OR_{ak}) = \frac{\min\limits_{d \in RD_{ak}} d + \max\limits_{d \in AD_{ak}} d}{2} \tag{3}$$

where $RD_{ak} \subset OR_{ak}$ is the set of demands that the proposer a has seen rejected and $AD_{ak} \subset OR_{ak}$ is the set of demands that the proposer a has seen accepted. The function thus uses two indicators to estimate the norm: the minimum of the rejected demands (RD_{ak}) and the maximum of the accepted demands (AD_{ak}). The rejection of a given demand indicates that the demand is higher than expected, thus the norm is lower than the minimum rejected demand. The acceptance of a given demand indicates that the demand is lower than expected (or perfect), thus the norm is higher (or equal) to the accepted demand. Equation (3) determines the norm as the average of these two indicators. By using

(3), we enable our agent to choose a demand following [5] and [10] theories on norms.

We extend the model provided by [16] to specify normDemand(OR_{ak}) for three edge cases: a case where there are no observed responses ($OR_{ak} = \emptyset$), a case with only observed rejections ($RD_{ak} = OR_{ak}$), and a case with only observed accepts ($AD_{ak} = OR_{ak}$):

- $OR_{ak} = \emptyset \rightarrow$ normDemand(OR_{ak}) is drawn from the normal distribution $N(561.8, 128.9)$, which is the normal distribution human demands follow in the empirical dataset we use of [3].[2]
- $RD_{ak} = OR_{ak} \rightarrow$ the agent averages between the minimum reject and halve the pie $((\min_{d \in RD_{ak}} d + 0.5P)/2)$.
- $AD_{ak} = OR_{ak} \rightarrow$ the agent averages between the maximum accept and the full pie $((\max_{d \in AD_{ak}} d + P)/2))$.

In [16], the agent determines a randomized demand as the norm in all these cases. By specifying these edge cases, we improve the SHA model aiming to better represent human behavior.

Agent with Values and Norms. This paper combines the value-based agent and the norm-based agent to model an agent that uses both values and norms. Values and norms are decision-making concepts the agent uses to decide what action is good and what action is bad [17]. [6] used values as more abstract ideals and norms as the more concrete context-dependent rules that follow from these ideals. Norms follow from values, but when agents copy the norms (but not values) from other agents the two can conflict. For example, one copies working over-hours although this is in conflict with its own values. This paper presents a model where an agent uses a weight (vw_a) to combine the best action based on values and the best action based on norms into the best action based on both values and norms.

For a proposer an action is defined as a demand $d \in \{0...P\} \subset \mathbb{Z}$. The demand d for an agent a in round k is determined by

$$d_{ak}(di_a, vw_a, OR_{ak}) = vw_a \times valueDemand(di_a) + (1 - vw_a) \times normDemand(OR_{ak}), \quad (4)$$

where di_a stands for the difference in importance an agent attributes to wealth versus the importance it attributes to fairness, $vw_a \in [0, 1]$ stands for the weight an agent attributes to its own values (versus the weight it attributes to norms) and OR_{ak} stands for the set of observed replies the agent has seen. Thus what an agent demands is the result of weighting the result of two functions described in [16]: the $valueDemand(di_a)$ (1) and the $normDemand(OR_{ak})$ (3).

[2] The norm that is drawn from the normal distribution is not used as input for the norm in subsequent rounds (i.e., the agent does not memorize it).

The above focuses on the demands of the SHA-proposer (SHA-P), but to simulate the UG we also need SHA-responders (SHA-R). The SHA-R is defined the same way as to the proposer model except that it determines a threshold $t \in [0, P]$ (instead of a demand d) and has to base its norms on a set of observed demands (instead of observed responses). The SHA-R uses the same function as the SHA-P (4) to determine the threshold. If the demand is higher than the threshold the responder rejects it, if the demand is lower than or equal to the threshold accepts it. The responder determines a threshold appropriate for its values analogue to the proposer, that is, by using (1) and (2). The responder determines the norm for the threshold by averaging over the observed demands (instead of the observed responses). The proposed demand is thus seen as a signal from the proposer that this is what the proposer considers 'normal'.[3] By using a threshold t based on its values and the norm (from observed demands), the responder uses values and norms to reject or accept the proposed demand.

Above we presented a model that uses both values and norms using two agent-specific variables: the difference in importance (di) and the value-norm weight (vw). The difference in importance specifies how much an agent values wealth over fairness and is normally distributed over agents $N(\mu_{di}, \sigma_{di})$, being μ_{di} the average value and σ_{di} the standard distribution of a normal distribution for di. The value-norm weight specifies how much an agent weighs values over norms and is normally distributed over agents $N(\mu_{vw}, \sigma_{vw})$. Both variables are constant over the different rounds. However, the $normDemand(OR_{ak})$ differs per round as the observed replies vary. To reproduce the demands and responses humans give, the parameters μ_{di}, σ_{di}, μ_{vw} and σ_{vw} need to be calibrated to the right settings.

2.3 Calibrating the SHA to Humans

We found that $\mu_{di} = 0.5$, $\sigma_{di} = 0.25$, $\mu_{vw} = -0.6$ and $\sigma_{vw} = 1.14$ produced simulated demands and responses that are closest to the empirical data on human demands and responses extracted from the dataset provided by [3]. This meta-study combines the data of 6 empirical studies to create a dataset of 5950 demands and replies made by humans in the same scenario as we describe in Sect. 2.1. We used five performance measures for which we compared the synthetic data on the SHA to the empirical data on real humans: average demand μ_d, standard deviation in demand σ_d, average acceptance rate μ_a, standard deviation in acceptance rate σ_a and the standard deviation in the demand solely based on values σ_{vd}.[4] The SHA is compared to humans on both round 1 and round 10. We used the following procedure to find the optimal parameter settings:

[3] To be exact, the proposed demand should be considered as what the proposer considers a 'normal' threshold. If it considered a higher threshold to be normal it would have demanded less, if it considered a lower threshold normal it would have demanded more.

[4] The standard deviation in the demand solely based on values σ_{vd} was added to ensure agents vary in what values they find important (di_a). The σ_{vd} for humans is postulated instead of extracted from empirical data.

1. Run simulations of the UG in Repast Simphony with different parameter settings ($\mu_{di} \in [-1, 1]$ and $\mu_{vw} \in [0, 1]$) and different random seeds ($r \in [1, 30]$) and obtain the resulting demands and acceptance rate (μ_d, σ_d, μ_a, σ_a);
2. For each run: average the performance measures with the same parameter settings, but different random seeds;
3. For each parameter setting: calculate the normalized root mean square error (NRMSE) between the simulated results and the human results;
4. Pick the parameter setting with a minimal NRMSE.

Table 1 compares the resulting demands and responses for the SHA (when NRMSE is minimized) and the empirical data on human results. We interpret these results as that our SHA can produce a distribution of demands and accepts that it is fairly close to that of humans (NRMSE = 11.0). The remainder of the paper uses these parameter settings to simulate human behavior based on values and norms.

Table 1. A comparison of the distribution of demands and responses between empirical data on humans and synthetic data on simulated human agents (SHAs) given the parameter settings with the best fit ($\mu_{di} = 0.5$, $\sigma_{di} = 0.25$, $\mu_{vw} = -0.6$ and $\sigma_{vw} = 1.14$).

Round	Source	Performance measures				
		μ_d	σ_d	μ_a	σ_a	σ_{vd}
1	Empirical (human)	561.8	128.9	0.806	0.40	128.9
	Synthetic (SHA)	557.9	91.1	0.876	0.29	109.2
10	Empirical (human)	584.2	98.66	0.868	0.34	122.5
	Synthetic (SHA)	646.8	90.89	0.923	0.23	109.2

3 Estimating Relative Preferences on Values and Norms

The conceptual model of the estimation process is presented in Fig. 1. The *Profiling Agent* (PA) is responsible to estimate the relative preference of a given SHA-P towards values and norms ($[\hat{di}_a, \hat{vw}_a]$), and whenever necessary, interact with the SHA-P to reduce ambiguity in the estimation.

The problem of estimating values and norms in the context of this work is translated to the estimation of di_a and vw_a, represented as $[\hat{di}_a, \hat{vw}_a]$. The estimation of relative preferences from behavior was assessed in different ways, such as via Inverse Reinforcement Learning (IRL) algorithms [1,12,14,18] and learning a utility function with influence diagrams [19]. In this work, we use the UG as a simple context (two values and one norm), aiming for clarity on understanding the conditions necessary to properly estimate the relative value

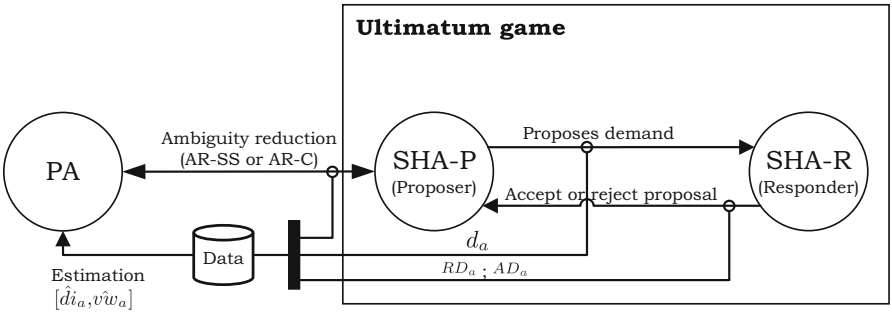

Fig. 1. Conceptual model.

and norm preference from behavior and, whenever necessary, how to reduce ambiguity.

In the context of the UG we can perform the estimation of the relative preferences by a exhaustive search process on the estimators, according to:

$$[\hat{di}_a, \hat{vw}_a] = \underset{di_a \in [-0.15;1.79]; vw_a \in [0;1]}{\arg\min} \frac{\sum_{k=1}^{m} \left| \hat{d}_{ak}(di_a, vw_a, OR_{ak}) - d_{ak} \right|}{k} \quad (5)$$

where $\hat{d_{ak}}$ is calculated by (4), and m represents the number of rounds from the UG analyzed in the estimation. In short, the estimators $\hat{di}_a$ and $\hat{vw}_a$ are defined as the preference set that minimizes the deviation between the observed demand (d_{ak}) and the demand that was calculated by an exhaustive search process ($\hat{d_{ak}}$). The range for di_a which has an effect on the demand by the SHA-P is $[-0.15; 1.79]$. With $vw_a \in [0; 1]$, and using a step of 0.01 to explore both variables, 19,392 evaluations of the fitness function are necessary for each estimation process.

3.1 Reducing Ambiguity

Ambiguity arises whenever the number of elements of $\hat{di}_a$ and $\hat{vw}_a$ is greater than 1. The PA will try to reduce this ambiguity by taking the role of the *responder* on the UG. This interaction might be interpreted as an elicitation process or, in simple, the PA will ask the SHA-P "questions". Different ways of "asking these questions" may produce higher or lower quality answers, and consequently a higher or lower change in the confidence on the estimation. We will explore two different approaches to how the PA interacts with the SHA-P. These interactions are a "side-game", i.e. changes in OR_a will not, by definition, influence future actions of the SHA-P.

Ambiguity Reduction by Exploring the Search Space (AR-SS). The approach to ambiguity reduction via exploring the search space ($AR-SS$) poses

the following "question" to the SHA: *What would your next demand be?*. Based on the demand ("answer") provided by the SHA-P, the PA will reject or accept it to explore the search space (Algorithm 1). This approach increases the search space, i.e. the distribution of OR_a in the dataset, by rejecting demands lower than rejected before and accepting demands higher than accepted before.

Algorithm 1: Ambiguity Reduction by exploring the Search Space ($AR-SS$)

input : $\hat{di}_a$, $\hat{vw}_a$, max_{int},RD_a,AD_a,OR_a, d_a, k
output: $\hat{di}_a$, $\hat{vw}_a$,RD_a,AD_a,OR_a, $count$

$n_{solutions} \leftarrow$number of solutions ($[\hat{di}_a, \hat{vw}_a]$)
$k \leftarrow k+1$
Calculate OR_{ak}
$count \leftarrow 0$
while $n_{solutions} > 1$ *AND* $count < max_{int}$ **do**
  Calculate $\hat{d}_{ak}(\hat{di}_a, \hat{vw}_a, OR_{ak})$ (4)
  Include OR_{ak} and $\hat{d}_{ak}$ to the data set
  Estimate $[\hat{di}_a,\hat{vw}_a]$ (5)
  $n_{solutions} \leftarrow$ number of solutions ($[\hat{di}_a, \hat{vw}_a]$)
  if $d_{ak} < min(RD_a)$ *OR* $\nexists$ RD_a **then**
    $RD_{ak} \leftarrow d_{ak}$
  else if $d_{ak} > max(AD_a)$ *OR* $\nexists$ AD_a **then**
    $AD_{ak} \leftarrow d_{ak}$
  $k \leftarrow k+1$
  Calculate $OR_{a,k}$
  $count \leftarrow count + 1$

Ambiguity Reduction via Counterfactuals (AR-C). Counterfactuals are mental representations of alternatives to events that have already occurred [24], frequently represented by conditional propositions related to questions in the "what if" form. Philosophical discussion on counterfactuals has been present for ages, including works from David Hume and John Stuart Mill [22], and it is considered an intrinsic element of causality. People use counterfactuals often in daily life to create alternatives to reality guided by rational principles.

Our approach to reduce ambiguity on the estimation of values and norms via counterfactual ($AR-C$) poses the following "question" to the SHA-P: *What would your next demand be if your opponent had accepted instead of rejected your proposal on round "x"?* or *What would your next demand be if your opponent had rejected instead of accepting your proposal on round "x"?*. As presented in Algorithm 2, the PA will 'ask' the 'question' related to the round that will lead to a broader search space in terms of the observed social norm (OR_a).

Algorithm 2: Ambiguity Reduction via Counterfactuals (AR-C)

input : $\hat{di}_a$, $\hat{vw}_a$, RD_a,AD_a,OR_a, d_a,k
output: $\hat{di}_a$, $\hat{vw}_a$,RD_a,AD_a,OR_a, *count*

$max_{int} \leftarrow k$
$count \leftarrow 0$
$n_{solutions} \leftarrow$number of solutions $([\hat{di}_a, \hat{vw}_a])$
while $n_{solutions} > 1$ *AND* $count < max_{int}$ **do**
  for $i = 1 : k$ **do**
    if $\exists$ AD_{ai} *(Proposal was* ***accepted*** *in round i)* **then**
      if $d_{ai} < min(RD_a)$ *OR* $\nexists$ RD_a **then**
        $RD_{ai} \leftarrow d_{ai}$
    else if $\exists$ RD_{ai} *(Proposal was* ***rejected*** *in round i)* **then**
      if $d_{ai} > max(AD_a)$ *OR* $\nexists$ AD_a **then**
        $AD_{ai} \leftarrow d_{ai}$
    Calculate OR_{ai}
    $score_i \leftarrow min(|OR_{ai} - OR_a|)$
  $k \leftarrow k + 1$
  $OR_{ak} \leftarrow OR_{az}$, where z is the iteration where *score* is maximum
  Calculate $\hat{d}_{ak}(\hat{di}_a, \hat{vw}_a, OR_{ak})$ (4)
  Include OR_{ak} and d_{new} to the data set
  Estimate $[\hat{di}_a, \hat{vw}_a]$ (5)
  $n_{solutions} \leftarrow$number of solutions $([\hat{di}_a, \hat{vw}_a])$
  $count \leftarrow count + 1$

4 Results

To test the methods presented in this paper we analyzed the behavior of agents acting as proponents (SHA-P) in 100 runs of the UG (each with 20 rounds), in terms of precision and confidence. The first will be evaluated by the root mean squared error (RMSE) between the observed demand (d) and the estimated demand ($\hat{d}$), while the latter by the number of elements in $[\hat{di}_a, \hat{vw}_a]$. A small number of elements represent high confidence (low ambiguity), while a large number of solutions represent low confidence (high ambiguity). It is guaranteed the number of elements of $[\hat{di}_a, \hat{vw}_a]$ is greater than zero, since we perform an exhaustive search in the complete range of di_a and vw_a. With these results, we aim to analyze the conditions under which the proposed methods can estimate preferences on values and norms in a precise and confident manner.

4.1 Estimation of Values and Norms

The precision in the estimation increased with the number of rounds used, given that an estimation is made by observing at least four rounds (Fig. 2(a)). Using less than four rounds the estimated values and norms ($[\hat{di}_a, \hat{vw}_a]$) predicted a

demand ($\hat{d}$) that is very close to the real demand[5], but most of these estimations will not be able to generalize among other contexts (i.e. other observed response - OR_a).

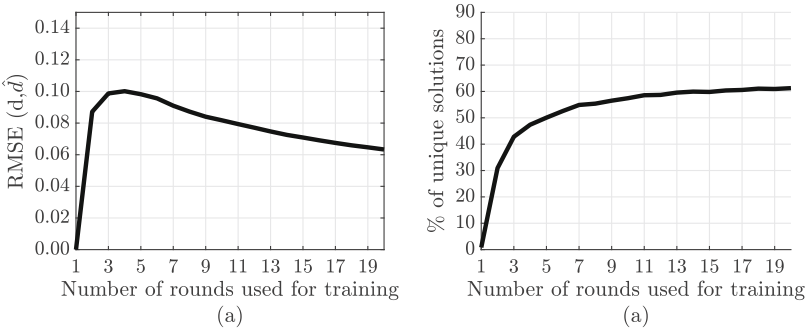

Fig. 2. (a) Precision of the estimation process. (b) Percentage of unique solutions for the estimation process.

When considering between one and four rounds the estimations were ambiguous: on average 69.2 different sets of estimations led to the same demand (Fig. 2(b)). Nevertheless, the confidence in the estimation increased with the number of rounds used. The steep curve reaching round four appears in both figures, showing a correlation between this initial precision in the estimation of the demand with the ambiguity on estimating preferences toward values and norms. Even with an increasing number of rounds observed, the percentual of unique solutions reached an average of only ca. 60%. These different estimations may lead to undesired properties of an agent that wants to act on behalf of the true values and norms of a given person.

4.2 Reducing Ambiguity

Both proposed methods were able to increase confidence in the estimation of preferences on values and norms ($[\hat{di}_a, \hat{vw}_a]$). Comparing Fig. 3(a) with Fig. 2(a), the RMSE between the calculated and observed demand are slightly higher, but still may be considered adequate in absolute values, given that $d \in \{0...1000\} \subset \mathbb{Z}$.

The methods $AR-SS$ and $AR-C$ did not increase confidence in the same manner, differing in their applicability. While $AR-SS$ was able to reach estimations with less ambiguity than $AR-C$ (Fig. 3(b)), it required in average of

[5] Given the deterministic model of the SHA, it might be expected that the RMSE should tend to zero. This is not the case because $[d, valueDemand, normDemand] \in \mathbb{Z}$, and therefore a rounding operator is used.

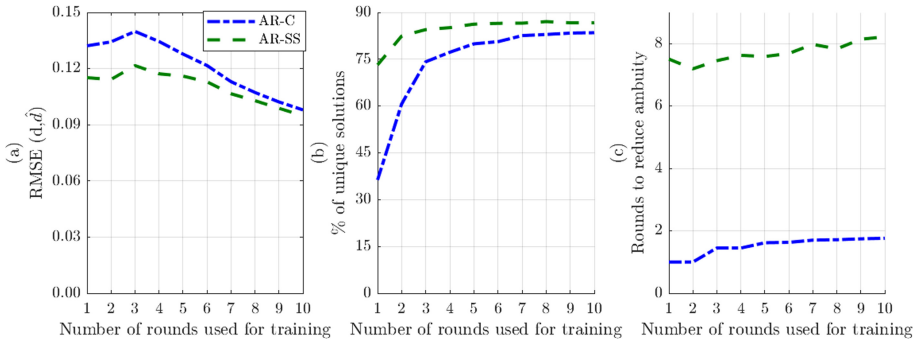

Fig. 3. Reducing ambiguity: (a) Precision; (b) Unique solutions; (c) Number of interactions made for reducing ambiguity.

4.9 more interactions with the user (Fig. 3(c))[6]. In 33.2% of the cases where the estimation was ambiguous, $AR-SS$ could provide a unique solution, while $AR-C$ provided it for 27.2% of the cases. Considering both the confidence and the number of interactions needed, $AR-C$ increased the number of unique solutions by 16.5% per interaction with the SHA-P, while $AR-SS$ increased it by 4.1% per interaction.

Table 2 summarizes the results when the estimation is performed using 10 rounds of observed data. We can see that precision varied only slightly, but the confidence in the estimation increased significantly when using any of the *ambiguity reduction* methods proposed. The columns related to the "Standard deviation" on Table 2 provide parameters to understand the dispersion of the search space[7] and of the estimations ($[\hat{di}_a, \hat{vw}_a]$). While the search space scatters with the ambiguity reduction methods, there is a reduction in the dispersion for the estimations. In other words, as the search space covered by OR_a increases, estimations become less ambiguous (and more gathered).

5 Discussion

AAs must be able to estimate human's relative preferences towards values and norms to align their behavior with them. This is not an easy task since a different set of preferences might lead to the same observed behavior (ambiguity problem). [18] demonstrated that 'normative assumptions' are needed to estimate an agent's reward function (in our case, values and norms). In this work,

[6] The x-axis in Fig. 3 relates to the number of rounds used during the initial estimation process, described in Sect. 4.1. The additional number of interactions performed by each method to improve the confidence in the estimations is not included in the x-axis but presented in Fig. 3(c).

[7] In our case OR_a: given that preference to values and norms are constant, demand is defined according to (4))

Table 2. Summary of the results for the estimation using 10 rounds, and for the subsequent ambiguity reduction methods.

Method	Precision	% Unique	Rounds on AR	Standard deviation		
				OR_a	$\hat{di}_a$	$\hat{vw}_a$
Estimation	0.082	57.4	–	36.1	0.021	0.008
AR-SS	0.095	86.7	8.2	42.6	0.011	0.003
AR-C	0.098	83.5	1.8	45.5	0.013	0.003

we proposed an ABM that uses both values and norms to account for these 'normative assumptions', and methods to estimate the SHA preferences and reduce the ambiguity in these estimations.

The first contribution, the proposed ABM, was built on previous works [15,17], and especially [16]. We extended the ABM works to use *both* values and norms for determining the actions of the agents. The model was calibrated to represent human behavior from empirical data. Future works can improve this model by specifying in more detail the use of norms and values, considering other values than wealth and fairness, constraining behavior by using deontic operators [16], and incorporating models that represent human bounded rationality. Furthermore, different application scenarios with a more complex view on value tensions e.g. non-zero sum game problems considering privacy and security can be modeled.

The second contribution of this work was focused on understanding to what extent the proposed methods were able to estimate the relative importance attributed to values and norms by a given agent. We show that even when considering 'normative' assumptions, represented by the PA's knowledge of the decision-making process of the SHA in a relatively simple context (the UG), ambiguity may still be present on estimating preferences for value alignment. To ignore this ambiguity might lead to great regret and unethical actions.

In the experiment performed we observed two general conditions under which an estimation of the preferences was possible, namely: heterogeneity on the observations and a sufficient number of observations. When observing at least four rounds of the UG, only 50 to 60% of the estimations were unambiguous. We proposed two methods for reducing ambiguity: one via exploring the search space ($AR-SS$) and one using counterfactuals ($AR-C$). The first approach, $AR-SS$, interacts with the SHA-P considering the last observed norm (OR_{ak}), which might be suitable for interacting with people in the real world due to the influence of short-term memory on decision making [7]. The latter approach, $AR-C$, uses counterfactuals, which are a part of causal reasoning. The results presented in the previous section show that targeting "imaginative" scenarios related to a specific round of the UG significantly increase the efficiency of the process. Nevertheless, it might be very demanding for a person to be able to go through this thought imaginary process and remind the precise social norm at a specific point in time. We can conclude that the $AR-SS$ method is more suitable

where there are no restrictions regarding the number of interactions with the user, while $AR-C$ can improve confidence in situations where a limited number of interactions is desired.

Both methods act with the assumption that the only way the PA can interact with the SHA is by taking the role of the responder in the UG. Going beyond this assumption, we also evaluated a different approach: the PA can directly define a value for the social norm (OR_{ak}). If we consider $OR_a \in [0; 1000]$, the ambiguity was almost eliminated: 97% of unique solutions on average, considering up to 20 interactions. If we consider a more realistic assumption $OR_{ak} \in [500; 1000]$ the results were, in terms of the percentage of unique solutions, between the levels found by $AR - SS$ and $AR - C$ when using less than 6 rounds for training, and slightly better than $AR - SS$ when using 6 or more rounds. Future works can evaluate this hypothesis. We suggest that such experiments be done either considering improved models of human memory and cognition process or in laboratory settings.

The limitations of this works include the impossibility of directly generalizing the findings and methods to other contexts, the assumption that values are stable, and the lack of testing of the approach with humans in realistic settings as well as in more complex settings.

6 Conclusion

This paper aimed to investigate to what extent an AA might be able to estimate the relative preferences attribute to human values and norms, including methods to reduce ambiguity. Insight into the use of models to support the estimation of values and norms was obtained during the discussions, mainly the need for heterogeneity on the observations and also ways to reduce ambiguity. Especially the use of counterfactuals via the $AR - C$ approach showed it can be of great value in terms of a trade-off between increasing efficiency and avoiding excessive interactions/questions with humans. We showed that even in a simple context, considering models that represent values and norms ('normative assumptions'), and using a exhaustive search process, ambiguity cannot be easily be avoided in estimation of preference on values and norms. To ignore this ambiguity might lead to great regret and misalignment between machine behavior and human values and norms.

Acknowledgements. This work was supported by the AiTech initiative of the Delft University of Technology.

References

1. Abbeel, P., Ng, A.Y.: Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the Twenty-First International Conference on Machine Learning (ICML). ACM (2004)
2. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D.: Concrete problems in AI safety. arXiv preprint arXiv:1606.06565 (2016)

3. Cooper, D.J., Dutcher, E.G.: The dynamics of responder behavior in ultimatum games: a meta-study. Exp. Econ. **14**(4), 519–546 (2011)
4. Cranefield, S., Winikoff, M., Dignum, V., Dignum, F.: No pizza for you: value-based plan selection in BDI agents. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), pp. 178–184 (2017)
5. Crawford, S.E.S., Ostrom, E.: A grammar of institutions. Polit. Sci. **89**(3), 582–600 (2007)
6. Dechesne, F., Di Tosto, G., Dignum, V., Dignum, F.: No smoking here: values, norms and culture in multi-agent systems. Artif. Intell. Law **21**(1), 79–107 (2013)
7. Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A.M., Nilsson, L.G.: The multifold relationship between memory and decision making: an individual-differences study. J. Exp. Psychol.: Learn. Mem. Cogn. **39**(5), 1344 (2013)
8. Dignum, V.: Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30371-6
9. Fehr, E., Fischbacher, U.: The nature of human altruism. Nature **425**(6960), 785–791 (2003)
10. Fishbein, M., Ajzen, I.: Predicting and Changing Behavior: The Reasoned Action Approach. Taylor & Francis Ltd, Milton Park (2011)
11. Güth, W., Schmittberger, R., Schwarze, B.: An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. **3**(4), 367–388 (1982)
12. Hadfield-Menell, D., Milli, S., Abbeel, P., Russell, S.J., Dragan, A.: Inverse reward design. In: Proceeding of the 31st Conference on Neural Information Processing Systems (NIPS), pp. 6765–6774 (2017)
13. Irving, G., Askell, A.: Ai safety needs social scientists. Distill **4**(2), e14 (2019)
14. Levine, S., Popovic, Z., Koltun, V.: Nonlinear inverse reinforcement learning with gaussian processes. In: Proceeding of the 31st Conference on Neural Information Processing Systems (NIPS), pp. 19–27 (2011)
15. Malle, B.F.: How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction. MIT Press, Cambridge (2006)
16. Mercuur, R., Dignum, V., Jonker, C.M., et al.: The value of values and norms in social simulation. J. Artif. Soc. Soc. Simul. **22**(1), 1–9 (2019)
17. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2018)
18. Mindermann, S., Armstrong, S.: Occam's razor is insufficient to infer the preferences of irrational agents. In: Conference on Neural Information Processing Systems (NIPS), pp. 5598–5609 (2018)
19. Nielsen, T.D., Jensen, F.V.: Learning a decision maker's utility function from (possibly) inconsistent behavior. Artif. Intell. **160**(1–2), 53–78 (2004)
20. Nouri, E., Georgila, K., Traum, D.: Culture-specific models of negotiation for virtual characters: multi-attribute decision-making based on culture-specific values. AI Soc. **32**(1), 51–63 (2014). https://doi.org/10.1007/s00146-014-0570-7
21. Oosterbeek, H., Sloof, R., Van De Kuilen, G.: Cultural differences in ultimatum game experiments: evidence from a meta-analysis. SSRN Electron. J. **8**(1), 171–188 (2001)
22. Pearl, J.: The seven tools of causal inference, with reflections on machine learning. Commun. ACM **62**(3), 54–60 (2019)
23. Van de Poel, I., et al.: Ethics, Technology, and Engineering: An Introduction. Wiley, Hoboken (2011)

24. Roese, N.J.: Counterfactual thinking. Psychol. Bull. **121**(1), 133 (1997)
25. Roth, A.E., Erev, I.: Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games Econ. Behav. **8**(1), 164–212 (1995)
26. Schwartz, S.H.: An overview of the Schwartz theory of basic values. Online Read. Psychol. Culture **2**, 1–20 (2012)
27. Soares, N., Fallenstein, B.: Agent foundations for aligning machine intelligence with human interests: a technical research agenda. In: Callaghan, V., Miller, J., Yampolskiy, R., Armstrong, S. (eds.) The Technological Singularity. TFC, pp. 103–125. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54033-6_5