Crowdsourcing Societal Tradeoffs

Vincent Conitzer
Dept. of Computer Science
Duke University
Durham, NC, USA
conitzer@cs.duke.edu

Markus Brill
Dept. of Computer Science
Duke University
Durham, NC, USA
brill@cs.duke.edu

Rupert Freeman
Dept. of Computer Science
Duke University
Durham, NC, USA
rupert@cs.duke.edu

ABSTRACT

It would be desirable if, as a society, we could reduce the amount of landfill trash we create, the amount of carbon dioxide we emit, the amount of forest we clear, etc. Since we cannot (or are in any case not willing to) simultaneously pursue all these objectives to their maximum extent, we must prioritize among them. Currently, this is done mostly in an ad-hoc manner, with people, companies, local governments, and other entities deciding on an individual basis which of these objectives to pursue, and to what extent.

A more systematic approach would be to set, at a global level, exact numerical tradeoffs: using one gallon of gasoline is as bad as creating x bags of landfill trash. Having such tradeoffs available would greatly facilitate decision making, and reduce inefficiencies resulting from inconsistent decisions across agents. But how could we arrive at a reasonable value for x?

In this paper, we argue that many techniques developed in the multiagent systems community, particularly those under economic paradigms, can be brought to bear on this question. We lay out our vision and discuss its relation to computational social choice, mechanism design, prediction markets, and related topics.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; J.4 [**Computer Applications**]: Social and Behavioral Sciences - Economics

General Terms

Algorithms, Economics, Theory

Keywords

social choice; judgment aggregation; information markets; mechanism design; crowdsourcing

1. INTRODUCTION

Society can agree, by and large, that certain activities (or by-products of activities) are undesirable: the emission of carbon dioxide, the creation of landfill trash, etc. This is a qualitative assessment. It is much more challenging to reach a quantitative assessment. For example, can we say that using one gallon of gasoline is just as bad for society as creating x bags of landfill trash? How would we arrive at a reasonable value of x? Such estimates would be extremely useful to policy makers as well as well-meaning institutions and individuals: it could help them make decisions that trade off multiple objectives (e.g., at what point should a car be taken off the road and replaced by a new, more fuel-efficient one?) and show them where to focus their efforts. It could also help in appropriately setting *Pigovian taxes*, which are taxes intended to discourage certain types of behavior (such as taxes on gasoline, alcohol, and tobacco). Finally, if agents act according to these tradeoffs, it will prevent inefficiencies resulting from inconsistent decision making across agents. For example, suppose agent 1 is in a position to significantly reduce landfill trash, but does not do so because she is more concerned about reducing gasoline consumption; whereas agent 2 is in a position to significantly reduce gasoline consumption, but does not do so because he is more concerned about reducing landfill trash. Both of them may prefer an outcome where they both perform their reductions, and achieving a consensus tradeoff first would likely guide them to do so.[1]

For certain types of tradeoff, an exact numerical value might be obtained purely from scientific evidence. For example, for two activities whose only downside is clearly the emission of a certain amount of carbon dioxide, we can simply measure these amounts and take the ratio. However, in most cases, including even the example above, it is hard to imagine that there is an objective fact of the matter as to what the correct value of x is. Rather, some collective subjective assessment, based at least in part on the preferences of members of the society, is necessary to establish a value for the tradeoff. At the same time, the assessment should still be informed by the relevant scientific evidence.

Our vision is to create a system that can credibly arrive at numerical values for societal tradeoffs, such as x above. The system should be flexible in that it generalizes across many similar questions. While our objective is narrowly defined, a successful solution seems to require the application of existing techniques from a variety of research areas, as well as the development of new techniques.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

[1]Some approaches have been proposed to allow the agents to directly reach a deal about their respective actions in contexts with externalities [10, 5, 6]. In contrast, we focus on deriving explicit global tradeoffs.

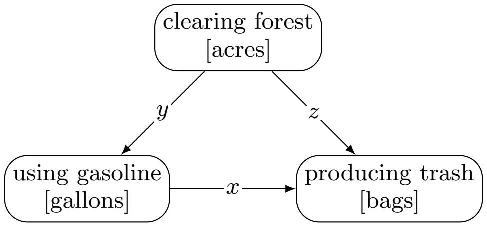

Figure 1: Weighted graph representing the numerical tradeoffs chosen by an individual voter. An arrow from activity A to activity B with weight w represents that one unit of A is considered as bad as w units of B. The tradeoff is consistent if $z = x \cdot y$.

2. RELATIONSHIPS TO EXISTING RESEARCH AREAS

This section is devoted to clarifying how existing research areas could be useful when building a system of the type described above. The next section (Section 3) highlights other issues that would need to be addressed.

2.1 Computational Social Choice

The approach we are considering is anchored first and foremost in the newly emerging field of computational social choice [2], which in turn is anchored in the broader field of social choice. Social choice is the study of how to aggregate the preferences of multiple agents into a collective choice or even a collective set of preferences. This is often done by having the agents *vote* over the alternatives.

It is not hard to see how voting could be applied to our motivating problem. For example, a natural approach is to let a collection of voters each report their estimate of x (their "vote"), and take the median of these numbers.[2] This *median voter rule* is well known in social choice theory [1, 12], but it will not suffice for our problem without some significant extensions. For one, voters in this context are likely to feel uncomfortable reporting a single number; they may prefer to give an interval of numbers within which they think the truth lies [9]. Another issue is that if we bring in a third undesirable activity to compare—say, clearing an acre of forest—then the tradeoffs should be consistent: if clearing an acre of forest is deemed as bad as using y gallons of gasoline, then it should be deemed as bad as creating $x \cdot y$ bags of landfill trash (see Figure 1). It turns out that achieving such consistency while maintaining the desirable properties of the median voter rule poses major challenges from a social-choice-theoretic viewpoint. In particular, the natural approach of aggregating the tradeoff for every *pair* of activities separately can lead to inconsistency (see Figure 2). This is reminiscent of similar challenges in the theory of *judgment aggregation* [7].

2.2 Prediction and Information Markets

The voting approach described above may suffice by itself when the voters constitute a body of experts who are all equally knowledgeable about every activity under consideration. But such a case will be the exception, rather than the rule. In general, each participant will be knowledgeable about only some of the activities under consideration. We would like our mechanism to naturally guide the participants to weigh in only (or at least more strongly) on the activities on which they are knowledgeable.

To achieve this, we can draw inspiration from *prediction markets* [16, 13]. A prediction market is used to obtain an estimate of the probability of a future event—e.g., what is the probability that Hillary Clinton will be elected the next president of the United States? About such a topic, too, we would expect some to be more knowledgeable than others. Indeed, prediction markets do not simply take the median of a set of estimates. Rather, in a prediction market, securities are traded. Such a security pays off (for example) \$1 if Clinton is elected president, and \$0 otherwise. The price at which these securities trade is then taken as the market's estimate of the probability that the event will occur. This naturally leads participants to trade on the events about which they feel they have superior knowledge. Is it possible to design a similar mechanism in our context, automatically guiding participants to evaluate activities (or pairs of activities) about which they have superior knowledge?

In fact, we would like to do more than this. Rather than having every participant act independently, we would like them to share information that they have, so that others can vote in a more informed manner. That way, we can still take into account the preferences of those who were initially uninformed. Potentially, we could even have a strict separation between experts, who teach voters about key facts that should inform the value of x, and voters, who vote on x. Previous research has been devoted to investigating how such information-providing participants can be integrated and rewarded in the context of prediction markets [4], and analogous techniques may be useful here.

Prediction markets typically rely on events eventually either taking place or not, so that we know which securities should pay out. By contrast, in our context, there may be no way to ever definitively verify who was right. However, contexts that are similar in this regard have been considered before. For example, when users rate a product, there may never be a final, definitive evalution of the product's quality that can be used to assess how accurate the ratings were. To address this, one approach to rewarding agents for accurate ratings is to see how well they predict future ratings by others—the *peer prediction method* [11] (see also work on the *Bayesian truth serum* [14] and much follow-up work). Something similar could be done in our context, namely, we could reward agents for having proposed tradeoff values that turned out to be close to the tradeoff values proposed by later agents.

2.3 Mechanism Design

A key issue is ensuring that each participant has every incentive to participate, and to participate honestly—as opposed to, say, voting for extreme values of x in order to obtain a final outcome that is better from her perspective. Indeed, the latter could happen if the final tradeoff value were obtained by taking the average rather than the median of the votes. But it is well known that voters have every incentive to vote truthfully when the median voter rule is used [12]. We should remember, though, that there are challenges in extending this rule in a consistent way to

[2]There are various social-choice-theoretic reasons for preferring taking the median to taking (say) the average, including limiting the influence of a single voter.

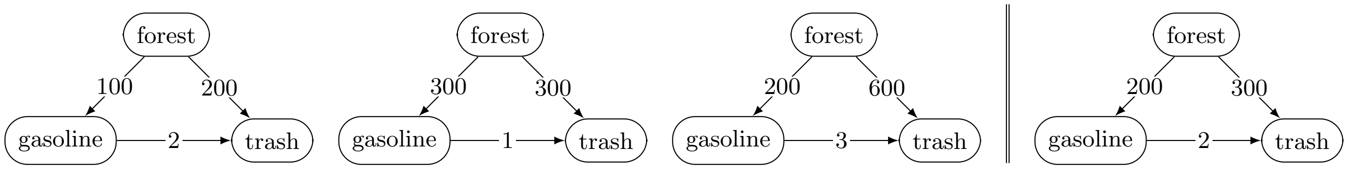

Figure 2: Example illustrating that the pairwise median rule can lead to inconsistent outcomes even when each individual voter is consistent. The left three graphs each illustrate the consistent preferences of a single voter, but the rightmost graph, which is inconsistent because $300 \neq 2 \cdot 200$, results from taking the median on each edge.

the case where we are comparing more than two activities. More broadly, the theory of *mechanism design* concerns how to design systems that result in good outcomes even when they are being used by strategic participants with potentially different interests, and it will clearly play an important role in this project.

If the system resulting from this project is to be used over the Internet, we face an additional challenge, which is that a single user may attempt to participate under multiple identities. Voting mechanisms, including the median voter rule, are quite vulnerable to this type of manipulation [3, 15], so we would need to develop techniques to address this, and/or monitor participants' identities to some degree. A number of techniques to address false-name manipulation in highly anonymous environments have already been considered in the literature; examples include making it difficult to obtain more than one account and thereby creating an effort cost to doing so, as well as investigating the social-network structure of the voters.

3. OTHER ISSUES TO ADDRESS

So far, we have discussed how we can bring ideas from several existing research areas in economic paradigms to bear on our motivating problem. In this section, we discuss several additional issues that should be considered.

3.1 Decomposition

When comparing two activities like using gasoline and creating landfill trash, rather than comparing them directly, it may make more sense to first break down their effects further. A potential approach could proceed as follows.

1. Identify the relevant *attributes* of these activities, where attributes correspond to societal goals or concerns. For example, gasoline use contributes to carbon dioxide emissions as well as to energy dependence, corresponding to two different attributes.

2. Determine how much the activity contributes to each attribute. In some cases this is directly measurable (the total carbon dioxide emitted from using one gallon of gasoline); in other cases this may itself be a matter of (collective) subjective judgment (the contribution of using a gallon of gasoline to energy dependence).

3. Finally, determine the tradeoffs among these attributes directly, in much the same way as was discussed previously.

This approach is likely to automatically guide the process to more thoroughly reasoned answers. It may also facilitate the process for the participants; for example, whereas it may be challenging to compare landfill trash creation and gasoline use directly, it may be easier to compare some of their attributes—e.g., the carbon dioxide emitted from the gasoline use vs. the methane emitted from the landfill. Finally, it may be the case that there are significantly fewer attributes than activities—e.g., many activities contribute to global warming but there are only few gases through which they do so—so that fewer comparisons between attributes will be needed. (On the other hand, we would still need to assess how much each activity contributes to each attribute.)

As already pointed out, both the process of evaluating the contribution of a given activity to a given attribute, and the process of evaluating the relative badness of two different attributes, constitute a (collective) subjective judgment problem of the form discussed earlier in this paper. So, it may appear that not much new is needed technically. However, one significant challenge is how to determine the list of attributes for a given activity. Specifying them *ex ante* runs the risk of missing an attribute that is important to some participants. Determining the attributes dynamically, for example by allowing participants to nominate new attributes on the fly, may be preferable, but would require a clear process to prevent the list from growing uncontrollably, having multiple copies of the same attribute, etc. This is particularly the case if we allow attributes to be broken down further (e.g., methane emissions into a global warming component and an explosion risk component).

How we break down activities into attributes has significant social-choice-theoretic implications. For example, suppose we break down gasoline use into two attributes, namely its effect on global warming and its effect on energy (in)dependence; and suppose we do not break down the creation of landfill trash any further. (If needed, we can create a single "landfill trash" attribute that is the only attribute of the "landfill trash" activity.) Moreover, suppose that it is uncontroversial (e.g., unanimously agreed) that gasoline contributes (in some units) 1 to global warming and 1 to energy dependence. Now suppose there are three voters.

- Voter 1 believes that 1 unit of global warming is as bad as 2 units of landfill trash, and 1 unit of energy dependence is as bad as 1 unit of landfill trash.

- Voter 2 believes that 1 unit of global warming is as bad as 1 unit of landfill trash, and 1 unit of energy dependence is as bad as 2 units of landfill trash.

- Voter 3 believes that 1 unit of global warming is as bad as 1 unit of landfill trash, and 1 unit of energy dependence is as bad as 1 unit of landfill trash.

Then, using the median rule in each case, we would conclude that 1 unit of global warming and 1 unit of energy dependence are each as bad as 1 unit of landfill trash; hence, 1 unit of gasoline is as bad as 2 units of landfill trash. On the other hand, voters 1 and 2 both feel that 1 unit of gasoline is as bad as 3 units of landfill trash (though for different reasons). Hence, if we had not broken down gasoline use further and rather compared it directly to landfill trash, then using the median rule we would have concluded that 1 unit of gasoline is as bad as 3 units of landfill trash. This is again a type of judgment aggregation paradox. It also illustrates that a party that gets to decide whether and how activities are broken down into attributes can have significant influence over the outcome; this is reminiscent of *control problems* in computational social choice [8]. All of this poses interesting questions for future research.

3.2 Objective vs. Subjective Tradeoffs

Especially when we consider decomposing activities into their relevant attributes, it becomes likely that in some cases, there is an objectively correct answer—e.g., how much carbon dioxide is emitted as a result of using a gallon of gasoline? In such cases, does it still make sense to have agents vote? Perhaps instead, it is preferable to classify some tradeoffs as "objective" and have a separate procedure for such cases. This, of course, would also require a procedure for classifying tradeoffs as subjective or objective. Ideally, we would find that our general procedure that involves agents voting (presumably with access to expert advice) does in fact result in the correct outcome in practice, when applied to objective tradeoffs. For one, this would obviate the need for special handling of objective tradeoffs. More importantly, many types of tradeoff are likely to border on the objective but retain a subjective component. If our general procedure tends to get the right answer in objective cases, it would increase our confidence in its outcomes on such borderline cases as well.

3.3 Local Tradeoffs

So far, we have considered settings where all participants arrive at a single tradeoff between a pair of activities. However, under some circumstances, it may make more sense to set tradeoffs locally. For example, it is not clear that clearing an acre of Amazon rainforest should be considered exactly as bad as clearing (say) one acre of the Belgrade Forest close to Istanbul. Of course, one could simply consider these to be two separate activities, but making too many activities distinct may result in a "thin markets" problem, where we have too few voters on each activity to be confident in our results. Can we dynamically decide when two closely related activities should be considered separate and when they should be merged into one? As another example, two countries may each want to use this system to set national policy. In this case, they may wish to arrive at separate conclusions even when the tradeoffs under consideration are clearly the same.

Additionally, locality does not need to refer to geographic locality. Instead, it could refer to locality in a social network. Particularly when the collective tradeoffs are used just to provide helpful guidance to individuals, rather than to set policy at a national or international level, it is not clear that everyone needs to obtain the same tradeoff as a result. For example, I may be interested to know how my friends feel about a certain tradeoff because I trust their ethical judgment; but I may have far less confidence in random members of the population providing me with (what I would consider to be) an ethically appropriate tradeoff. Under such circumstances, it may make more sense for *my* recommended tradeoff to be given by the median of my friends' votes.

Of course, setting tradeoffs locally is likely to result in the same inefficiencies discussed in the introduction (with the exception of the case where the activities are really distinct depending on the location in which they are done). Still, one may hope that setting tradeoffs locally will still result in decisions that are at least *more* consistent and efficient than they would be if they were made separately.

3.4 Selecting Application Domains

The directions discussed above would establish the fundamental theory needed for the project. However, to build a real, deployed system, much additional work would be needed, ranging from decisions about who gets to participate in the mechanism (and in what role) to the many design decisions faced in building a usable system. While the long-term goal is to develop general tools that can be used across a variety of domains, it is nevertheless likely that an initial system will be more successful if it focuses on a restricted domain—say, environmental issues. This will allow us to tailor the system appropriately; doing so effectively may require recruiting a domain expert for assistance.

4. CONCLUSION

Overall, we aim to create the theory and tools to better address the motivating problem, ideally culminating in a usable (e.g., web-based) system. We are currently (Spring 2015) teaching a special course at Duke University on this topic, thereby involving a broader team and providing them with the background knowledge needed to move the project further along to achieving its goals. We believe that this is an exciting opportunity to bring techniques developed in computational social choice and related areas to bear on important real problems.

Acknowledgments

We were supported by NSF and ARO under grants CCF-1101659, IIS-0953756, CCF-1337215, W911NF-12-1-0550, and W911NF-11-1-0332, and by a Feodor Lynen research fellowship of the Alexander von Humboldt Foundation.

REFERENCES

[1] D. Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

[2] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiß, editor, *Multiagent Systems*, chapter 6, pages 213–283. MIT Press, 2nd edition, 2013.

[3] V. Conitzer. Anonymity-proof voting rules. In *Proceedings of the Fourth Workshop on Internet and Network Economics (WINE)*, pages 295–306, Shanghai, China, 2008.

[4] V. Conitzer. Prediction markets, mechanism design, and cooperative game theory. In *Proceedings of the 25th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 101–108, Montreal, Canada, 2009.

[5] V. Conitzer and T. Sandholm. Expressive markets for donating to charities. *Artificial Intelligence*, 175(7–8):1251–1271, May 2011.

[6] V. Conitzer and T. Sandholm. Computing optimal outcomes under an expressive representation of settings with externalities. *J. Comput. Syst. Sci.*, 78(1):2–14, 2012.

[7] U. Endriss. Judgment aggregation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 17. Cambridge University Press, 2015.

[8] P. Faliszewski and J. Rothe. Control and bribery in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 7. Cambridge University Press, 2015.

[9] J. Farfel and V. Conitzer. Aggregating value ranges: preference elicitation and truthfulness. *Autonomous Agents and Multi-Agent Systems*, 22(1):127–150, 2011. Special Issue on Computational Social Choice.

[10] A. Ghosh and M. Mahdian. Charity auctions on social networks. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1019–1028, 2008.

[11] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[12] H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[13] D. Pennock and R. Sami. Computational aspects of prediction markets. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 26. Cambridge University Press, 2007.

[14] D. Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

[15] T. Todo, A. Iwasaki, and M. Yokoo. False-name-proof mechanism design without money. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 651–658, Taipei, Taiwan, 2011.

[16] J. Wolfers and E. Zitzewitz. Prediction Markets. *The Journal of Economic Perspectives*, 18(2):107–126, 2004.