Validation of Proxy Random Utility Models for Adaptive Adversaries

Richard S. John*[a] and Heather Rosoff[b]
[a] Department of Psychology, University of Southern California, Los Angeles, California, USA
[b] Sol Price School of Public Policy, University of Southern California, Los Angeles, California, USA

Abstract: We report two validation studies comparing MAU models for two different politically active non-profit organizations that utilize civil disobedience to achieve political objectives. In both cases, we constructed an objectives hierarchy and MAU model using adversary values experts (AVEs) who have access to publicly available information about the organizations' motives, objectives, and beliefs, but no direct contact with organization stakeholders or representatives. We then independently compare these MAU model parameters and constructed preferences to those based on direct assessment from a representative of the organization. The proxy MAU models provide an "averaged" utility model across a diverse organization with varying perspectives. We compare these "averaged" representations of the organizations' objectives, trade-offs, risk attitudes, and beliefs about consequence impacts with those of individual organization representatives with a particular perspective. In both cases, we demonstrate moderate convergence between the proxy model and the model assessed by direct contact with a representative of the organization. Overall we find moderate agreement between the proxy model and the stakeholder model, with some notable discrepancies. Most of these discrepancies can be attributable to unstated or understated objectives in the published materials of the groups.

Keywords: Model validation, adaptive adversary, multi-attribute utility, expert elicitation.

1. INTRODUCTION

Game theoretic approaches to modeling adaptive adversaries require accurate representation of adversary preferences. Early models relied on a zero-sum assumption; attacker gains equal defender losses, and attacker losses equal defender gains. Such a zero-sum assumption is an oversimplification that is likely to result in misspecification of attacker preferences [1,2,3]. An accurate representation of attacker preferences requires that attacker beliefs and values be assessed and quantified. Multi-attribute utility (MAU) models provide a framework for representing adversary concerns, beliefs regarding attack consequences, risk attitudes, and value trade-offs. MAU modeling generally requires access to a decision maker in order to elicit objectives and assess model parameters.

Most adversaries are not available for or willing to allow for direct elicitation which is required to construct an MAU model. Such adversaries have a strong interest in countering or foiling others; examples range from criminal organizations, terrorist organizations, corporations seeking to gain market advantage, political organizations seeking to promote their views and hindering rivals from making progress, and sports rivalries. In such cases, it is necessary to construct a representation of preferences using information that is known about adversary motivations, objectives, and beliefs. Such information includes a variety of sources, including past adversary behavior, public statements by the adversary, adversary web sites, and intelligence. An adversary objectives hierarchy and MAU model based on this information can be constructed by proxy, using judgments from an adversary values expert (AVE).

The construction of value models by proxy raises the question of whether such models can accurately capture adversary preferences using only secondary and tertiary sources. There is no published research to date on the validity of utility models constructed by proxy. In this paper, we report two validation studies comparing MAU models for two different politically active non-profit organizations that utilize civil disobedience to achieve political objectives. In both cases, we constructed an objectives hierarchy and MAU model using AVEs who have access to publicly available information

*email: richardj@usc.edu

about the organizations' motives, objectives, and beliefs, but no direct contact with organization stakeholders or representatives. We then independently compare these MAU model parameters and constructed preferences to those based on direct assessment from a representative of the organization. The proxy MAU models provide an "averaged" utility model across a diverse organization with varying perspectives. We compare these "averaged" representations of the organizations' objectives, trade-offs, risk attitudes, and beliefs about consequence impacts with those of individual organization representatives with a particular perspective. In both cases, we demonstrate moderate convergence between the proxy model and the model assessed by direct contact with a representative of the organization.

We use these two case studies to explore possible advantages of a proxy utility function compared to one directly assessed from the decision maker. Most adversary groups are composed of a variety of different stakeholders, each with a particular set of motivations and priorities. By accessing all available information about the group, one can include a broader range of perspectives in the adversary utility model than might be available from an interview with a single decision maker or stakeholder. Another possible advantage of using an AVE relates to the well known problem of eliciting a complete set of objectives from a decision maker [4,5]. Past research has demonstrated that decision makers may omit almost half of the important objectives when interviewed directly about their concerns and objectives. This may be due to the common use of heuristic tools in which the decision maker focuses on only the most salient or most important attributes. A knowledgeable AVE, or proxy decision maker, has the advantage of dispassion, and may be able to delineate a more comprehensive list of value relevant objectives.

In both cases, we obtain group stakeholder models from a single stakeholder. As one might expect, there are intra-group differences among stakeholders for any politically active group, and our two case study groups are no exception. In the end, we are able to evaluate convergent validity between our proxy utility model developed by an independent AVE with no direct contact with any group stakeholder, and the model developed from a single stakeholder. In the end, it is not clear which model is closer to the "centroid" of group values, as the stakeholder represents one point of view, while the AVE model represents the public face of the organization.

2. OVERVIEW OF ADVERSARY UTILITY MODELING

We utilize a Stackelberg game formulation in which the defender plays the role of the leader and the attacker (adversary) plays the role of the follower. The adversary is adaptive, in the sense that he is able to observe the defender's probabilistic strategy before deciding on an attack strategy. The defender must consider the adversary's motivations and values in order to select a strategy that mitigates the adversary's ability to adapt. Figure 1 displays an overview of the Stackelberg game using an influence diagram representation. The defender's decision node is represented in blue and the adversary's decision node is represented in red. The focus of this paper is on the representation of adversary values and beliefs, including consequence uncertainties, trade-offs, and risk-attitudes. These are colored purple, to indicate that the blue defender has explicitly modeled the red attacker's values and beliefs, and has incorporated them as uncertainties in her own model. For the defender to select a strategy, she must also know her own values and beliefs, represented in blue in this diagram. Our focus, however, is on validating defender models of the attacker's values and beliefs.

Unfortunately, defenders often make the simplifying assumption that they are playing a zero-sum game, in which the attacker's values are the inverse of their own values, and that the attackers' beliefs are identical to their own beliefs. These assumptions are hardly ever justified, and in fact are generally not even a good approximation of the attacker's values and beliefs. A better approach is for the defender to explicitly model the attacker's values and beliefs, using input from all available sources. Adversary beliefs and values are often described in detail in published writings and on web-sites. We have coined the term adversary values expert (AVE) for those intelligence experts who study the adversary's values and beliefs. Our approach is to identify an AVE for a particular adversary, and to

Figure 1. Influence Diagram Representation of Value Focused Adversary Random Utility Model

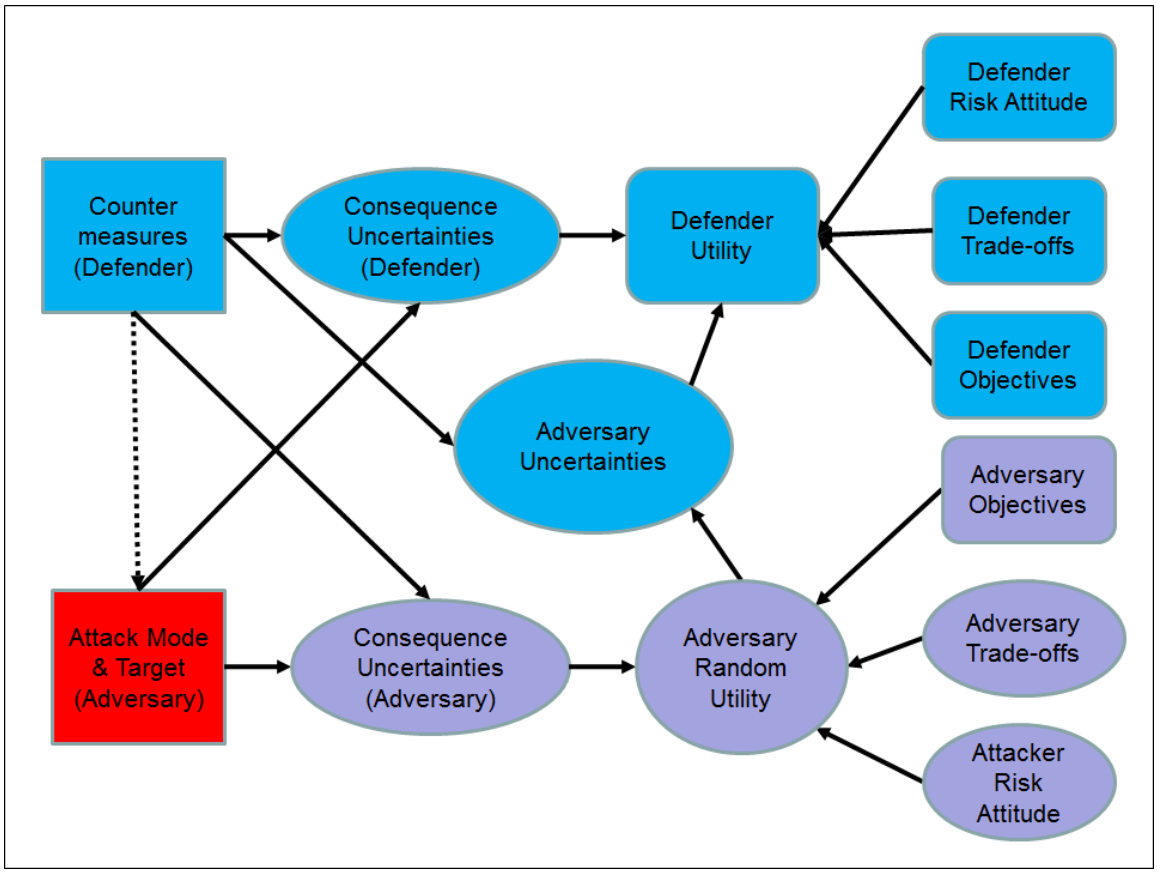
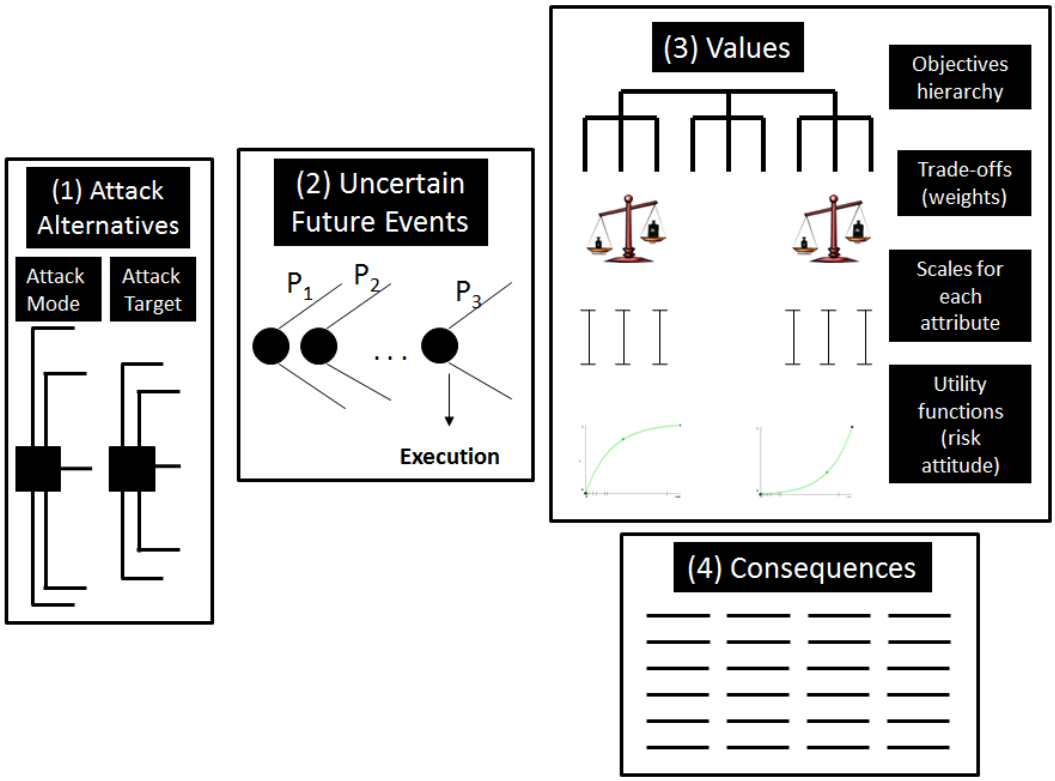


Figure 2. Schematic Overview of Adversary Value Model Elicitation

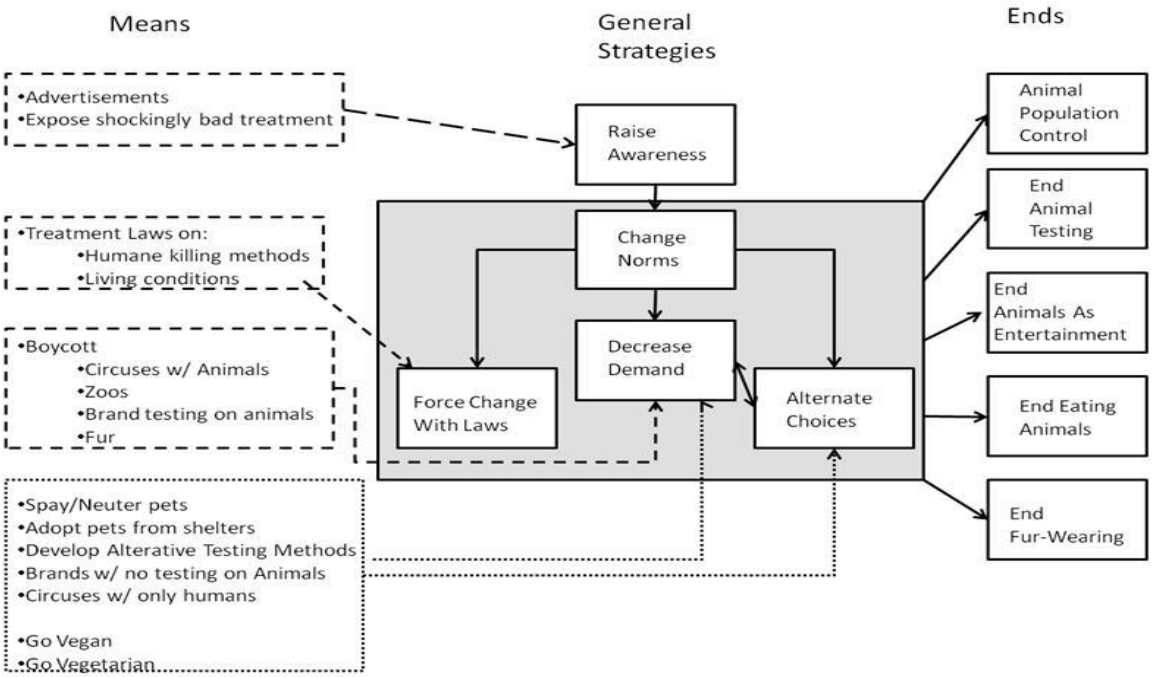


treat the AVE as a proxy decision maker for the adversary. We then construct a multi-attribute utility model for the adversary, using inputs from the AVE. Figure 2 displays a schematic overview of the MAU model and the elicitation process. The AVE identifies attack alternatives, which can be in the form of modes of attack, or particular attack targets, or timing of an attack, etc. The AVE also identifies relevant exogenous uncertainties that might affect the attack, such as whether the attacker can obtain required resources, or whether the attack will be interdicted. The AVE provides the fundamental objectives for the attacker. The AVE also constructs scales for mapping performance of attack alternatives on each objective. Standard MAU elicitation techniques are used to elicit the AVE's trade-offs (weights) among conflicting objectives and assess attitude toward risk for each attribute scale. Finally, the AVE provides the score matrix describing the performance of each alternative on the defined attribute scales; these may be either point estimates or distributions, reflecting uncertainty in the consequences of an attack.

3. CASE STUDY 1

The first validation study involved an animal rights group (ARG). We recruited a research assistant who was familiar with ARG 1 to become an AVE, learning everything about the ARG using published writings and internet sites. The AVE was not allowed to talk to ARG 1 stakeholders. The AVE worked independently, and did not take part in the stakeholder elicitation. In thinking about the ARG's fundamental objectives, it was useful to construct a means-ends objectives network, displayed in Figure 3. This representation allows the AVE to separate out means and ends objectives for the ARG. A fundamental objectives hierarchy for the ARG is displayed in Figure 4. The overall objective of improving animal rights is divided into three groups of objectives: maximize organizational power, minimize cruelty to animals, and maximize public perception of the group. Within these 3 groups of objectives, a total of ten fundamental objectives were identified by the AVE.

Figure 3. Adversary Means-Ends Objectives Network for ARG 1



Completely independent of the AVE's model construction, a parallel assessment was carried out by one of the authors with a stakeholder (employee) of the ARG. The stakeholder focused on the primary objective of the ARG to change the hearts and minds of the public with respect to animal consumption. The stakeholder identified six fundamental objectives for the ARG, and these are displayed in Figure 5. Despite the use of different terminology, it is evident that the AVE did capture the identical concerns expressed by the ARG stakeholder.

Figure 4. AVE Fundamental Objectives Hierarchy for ARG 1

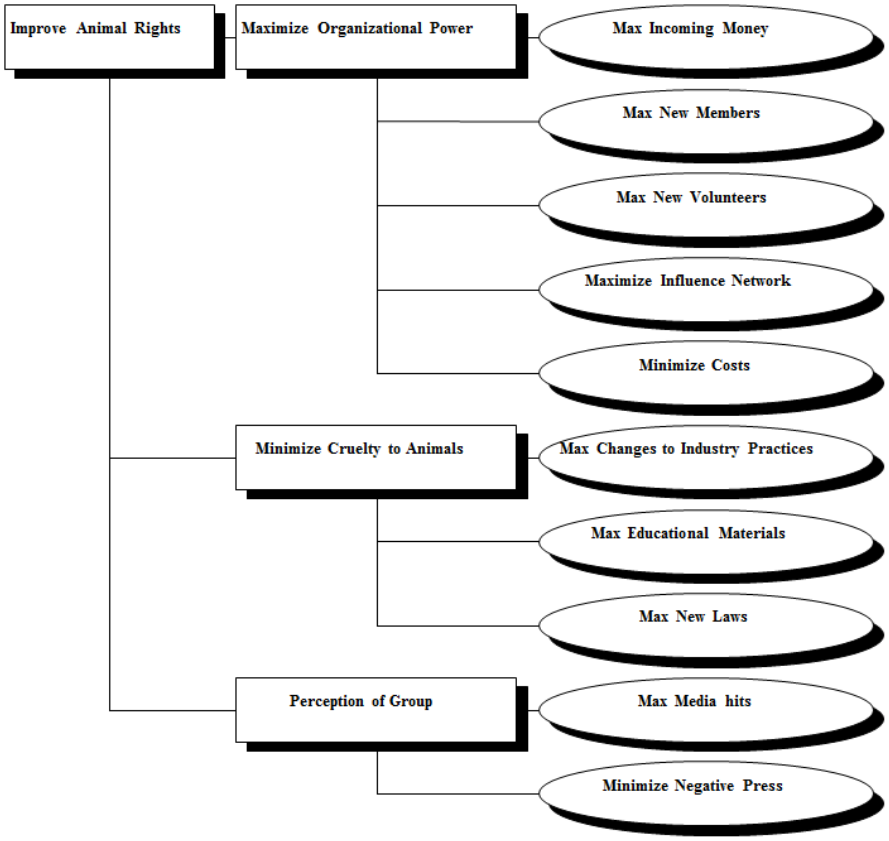
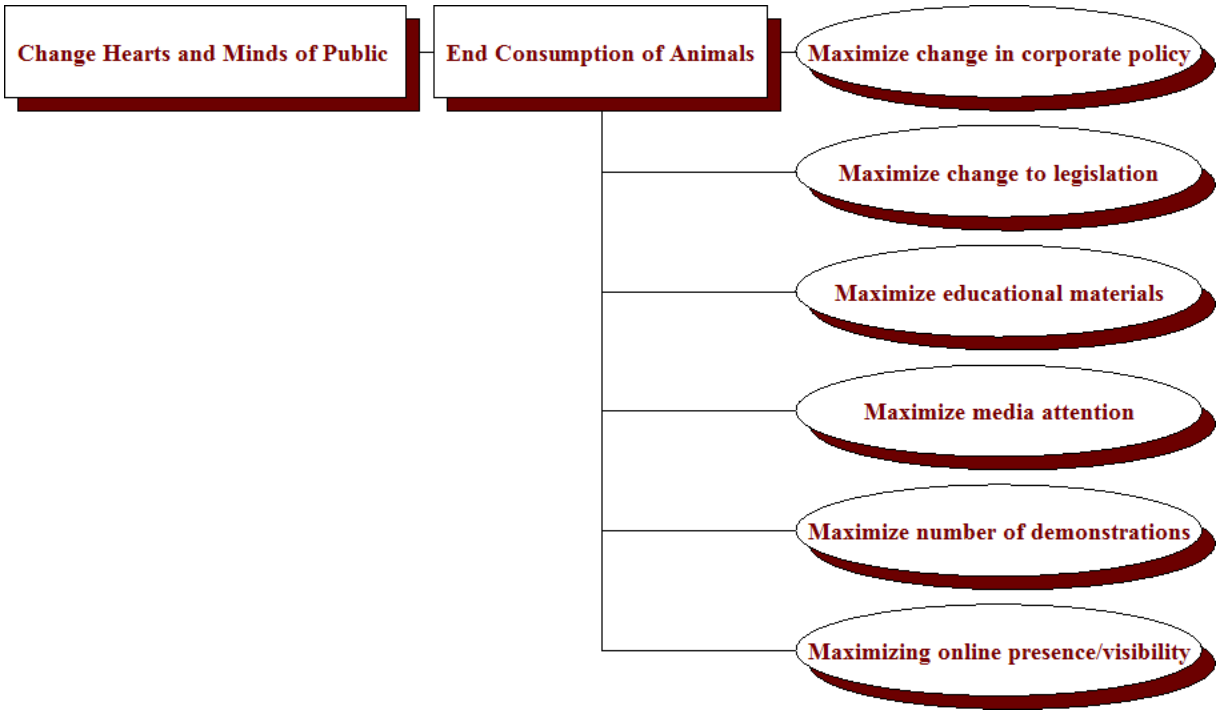


Figure 5. Stakeholder Fundamental Objectives Hierarchy for ARG 1



The various attack alternatives identified by the AVE and the ARG stakeholder are displayed in Table 1. There is a great deal of overlap (undercover investigations, ad campaigns, corporate campaigns), the ARG stakeholder did identify some attack alternatives not suggested by the AVE.

Table 1. AVE and Stakeholder Campaign Alternatives for ARG 1

Expert Campaign Alternatives	Stakeholder Campaign Alternatives
Undercover investigation (Animal Experiment Lab)	Demonstrations
New campaign (KFC)	Undercover investigations
New ad campaign	Online ad campaign
Publicity stunt	Print ad campaign
Purchase stock in company (McDonalds)	Corporate campaign
	Publicity stunt
	Policy enforcement

Both AVE and ARG stakeholder attribute scales are displayed in Table 2. Interestingly, all but one of the AVE's scales are based on percent changes from the status quo. The ARG stakeholder's scales are all natural scales based on observable performance metrics.

Table 2. AVE and Stakeholder Attribute Scales for ARG 1

EXPERT Attributes	Worst	Best	CE
Maximize changes to industry	-1	3	1.2
Maximize educational materials	-10%	30%	15.0
Maximize incoming money	-10%	30%	9.5
Maximize media hits	-10%	30%	15.5
Maximize new laws	-10%	30%	4.5
Maximize new members	-10%	30%	11.0
Maximize new volunteers	-10%	30%	10.5
Maximize influence network	-10%	30%	15.5
Minimize costs	-10%	30%	14.0
Minimize negative press	-10%	30%	17.5

STAKEHOLDER Attributes	Worst	Best	CE
Maximize media attention	0	15	10
Maximize online presence/visibility	1,000	40,000	15,000
Maximize educational materials	0	1,600,000	500,000
Maximize change in corporate policy	0	26	13
Maximize change in legislation	0	6	2
Maximize number of demonstrations	0	20	5

Trade-off parameters (weights) for both the AVE and ARG stakeholder models are presented in Table 3. Media attention does receive the greatest weight for both, but the ARG stakeholder (41%) puts over twice as much weight on media attention as the AVE (18%).

Table 3. AVE and Stakeholder Weights for ARG 1

ARG 1 Expert Swing Weights	
Max media attention	17.9
Max new laws	16.1
Max changes to industry practices	14.3
Max media hits	12.5
Max new volunteers	10.7
Max incoming money	8.9
Max educational materials	8
Max influence network	5.4
Min cost	4.5
Min negative press	1.8

ARG 1 Stakeholder Swing Weights	
Max media attention	40.8
Max online presence/visibility	24.2
Max educational materials	15.8
Max change in corporate policy	10.3
Max change to legislation	6.1
Max number of demonstrations	2.8

An additive MAU model was used to calculate expected utilities for both the AVE's model and the ARG stakeholder's model. For the AVE's model, uncertainties in the AVE's score matrix and uncertainties in both trade-offs and risk attitudes (single attribute utility functions) required use of Monte Carlo simulation to obtain expected utilities. For the ARG stakeholder, point estimates were obtained (independently) for all inputs, including the score matrix, weights, and utility function parameters. Table 4 presents expected utilities for alternatives using both the AVE's model and the ARG stakeholder model. There is strong agreement the four common alternatives, Pearson r = 0.81. Interestingly, undercover investigations scored highest in the AVE model and a close 2nd in the ARG stakeholder model. Note that demonstrations scored highest in the ARG stakeholder model, which is very similar to the publicity stunt option identified by the AVE. We suspect that some of the discrepancies between the two models are attributable to terminology rather than to differences in preference. However, there are some notable discrepancies. While corporate campaigns and print advertisement scored quite high in the AVE model, both scored quite low in the ARG stakeholder model.

4. CASE STUDY 2

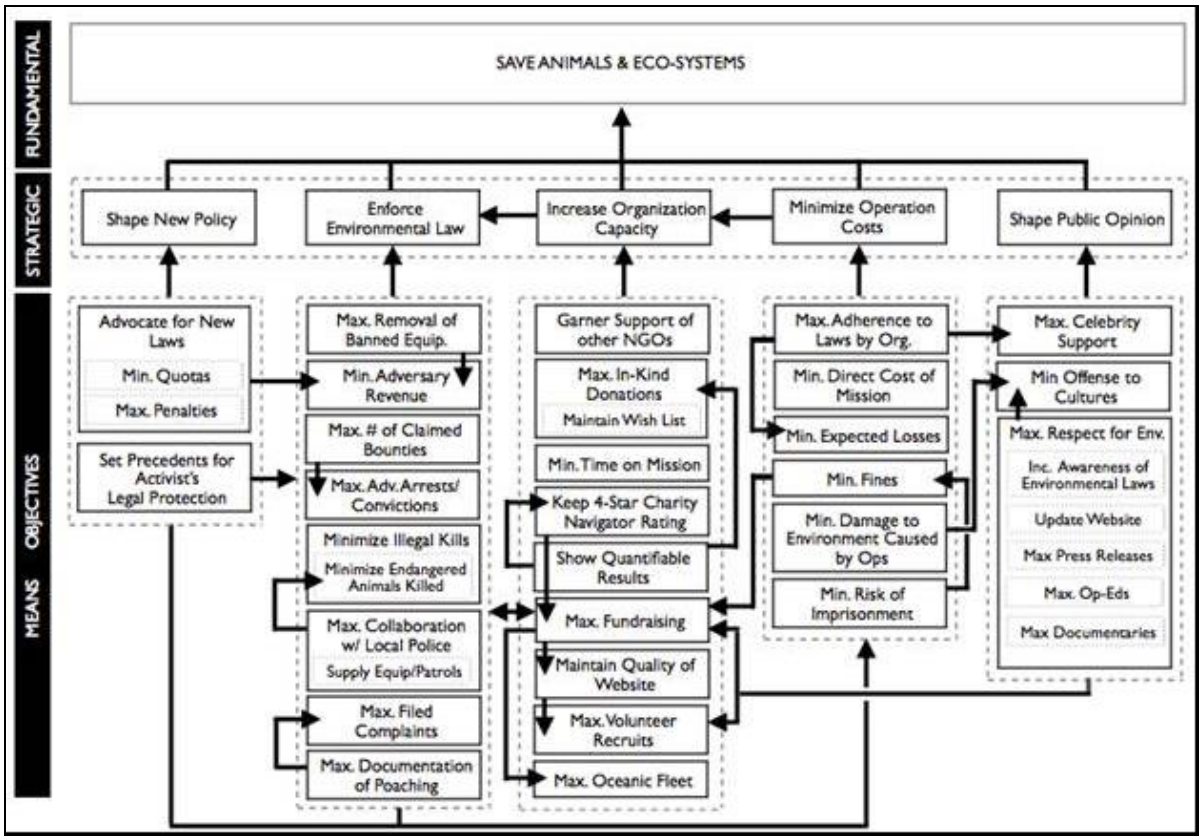
The second validation study involved a very different type of animal rights group and a different AVE. Again, the AVE was a research assistant who was instructed to learn everything about the group (ARG

2), using published writings and web sites, but without consulting with ARG 2 stakeholders. The AVE first constructed a means-ends objectives network for ARG 2, displayed in Figure 6

Table 4. AVE and Stakeholder Utilities for ARG 1 Campaign Alternatives

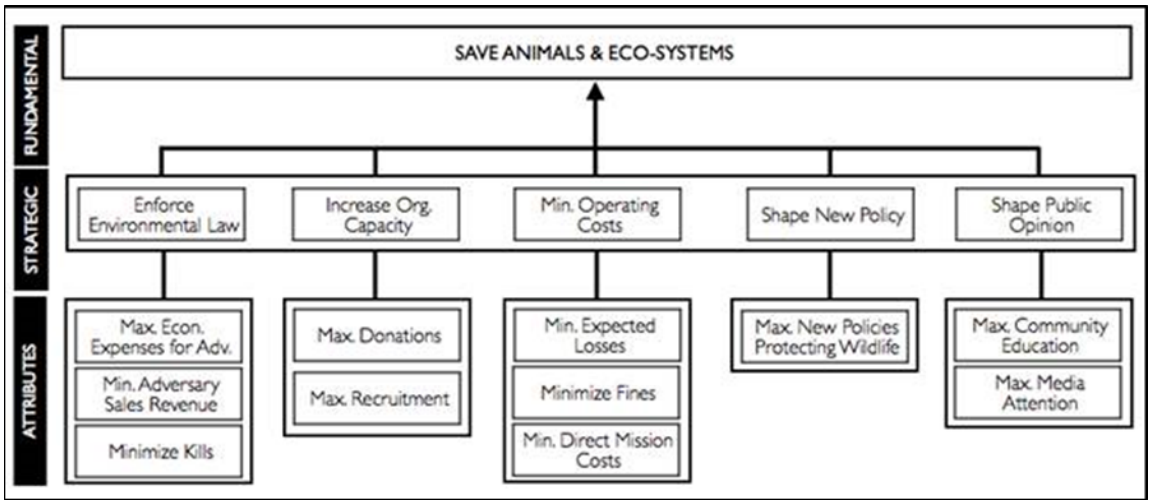
	E(U) Campaign	
	Expert	Stakeholder
Undercover investigation	0.56	0.33
Corporate campaign	0.55	0.14
Print advertisement	0.52	0.16
Publicity stunt	0.44	0.04
Purchase stock	0.37	
Demonstrations		0.37
Online advertisement		0.26
Policy enforcement		0.04

Figure 6. Adversary Means-Ends Objectives Network for ARG 2



From this means-ends network, the AVE was able to identify five fundamental objectives for ARG 2. These objectives, as well as the eleven attribute scales, are presented in Figure 7.

Figure 7. AVE Fundamental Objectives Hierarchy for ARG 2



This validation study was conducted somewhat differently from the first, in that the ARG stakeholder was provided the alternatives and model structure (fundamental objectives and attributes and score matrix) identified by the AVE, and was asked only to provide a complete set of weights defining his trade-offs among the conflicting objectives and attributes. Seven alternative campaign alternatives for ARG 2 were identified, and are displayed in Table 5.

Table 5. AVE Campaign Alternatives for ARG 2

Campaign Alternatives
Defending Pilot Whales in the Faroe Islands (North Atlantic Ocean).
Defending Dolphins in Taiji, Japan (Pacific Ocean).
Defending Blue Fin Tuna (Mediterranean Ocean).
Defending Sharks, Turtles & Eco-system within the Galapagos Islands (Ecuador).
Defending Harp Seals (Canada).
Defending Cape Fur Seals (Namibia).
Defending Whales (Antarctic Ocean)

The eleven attribute scales are displayed in Table 6. With the exception of fines, all are defined in terms of percentage of the status quo. As can be seen in the last two columns, the risk attitudes, defined by the certainty equivalents for 50-50 gambles between the worst and best, are quite different for the AVE (expert) and the ARG 2 stakeholder (SH). The AVE also scored each of the seven alternatives on all eleven attribute scales. Unlike the first case study, the AVE provided point estimates only (medians), thus there was no uncertainty information obtained for the score matrix. All 77 point estimates for the 7 alternatives by 11 attribute matrix is presented in Table 7.

Table 6. AVE Attribute Scales for ARG 2

Attribute	Worst	Best	Expert CE	SH CE
Min fines	$1 million	0	$200,000	0
Max economic impact for adversary	0%	1000%	800%	90%
Max new policies protecting wildlife	0%	100%	80%	50%
Max community education	0%	200%	80%	50%
Min adversary (i.e. poachers) revenue	0%	100%	80%	90%
Max media attention	0%	100%	60%	100%
Max donations	0%	100%	50%	38%
Max recruitment	0%	100%	40%	38%
Min illegal kills	100%	0%	20%	90%
Min direct costs of mission	100%	0%	20%	0%
Min expected losses	500%	0%	20%	0%

Table 7. AVE Alternatives by Attributes Score Matrix for ARG 2

Alternatives Matrix (medians)	SHAPE NEW POLICY	SHAPE PUBLIC OPINION		INCREASE ORG CAPACITY	
	Maximize New Policies Protecting Wildlife	Maximize Comm. Ed.	Maximize Media Attention	Maximize Recruitment	Maximize Donations
Defending Dolphins in Faeros	25%	60%	20%	50%	60%
Defending Dolphins in Taiji, Japan	5%	40%	70%	50%	70%
Defending Bluefin Tuna, Mediterranean	40%	30%	30%	40%	70%
Defending Sharks/Turtles in Galapagos	50%	8%	8%	5%	10%
Defending Seals in Canada	5%	35%	45%	30%	30%
Defending Cape Fur Seals in Namibia	75%	65%	50%	10%	30%
Whales (Antarctic Ocean)	5%	85%	90%	85%	80%

Alternatives Matrix (medians)	ENFORCE ENVIRONMENTAL LAW			MINIMIZE OPERATION COSTS		
	Minimize Illegal Kills	Maximize Economic Impact for Adversary (i.e.	Minimize Adversary (i.e. Poacher) Revenue	Minimize Fines	Minimize Direct Cost of Mission	Minimize Expected Losses
Defending Dolphins in Faeros	50	100%	10%	6,000	70%	50%
Defending Dolphins in Taiji, Japan	90	30%	5%	14,000	80%	60%
Defending Bluefin Tuna, Mediterranean	90	10%	5%	750,000	80%	30%
Defending Sharks/Turtles in Galapagos	70	25%	30%	500	40%	10%
Defending Seals in Canada	40	10%	15%	50,000	35%	50%
Defending Cape Fur Seals in Namibia	90	10%	5%	20,000	25%	70%
Whales (Antarctic Ocean)	30	600%	30%	5,000	95%	70%

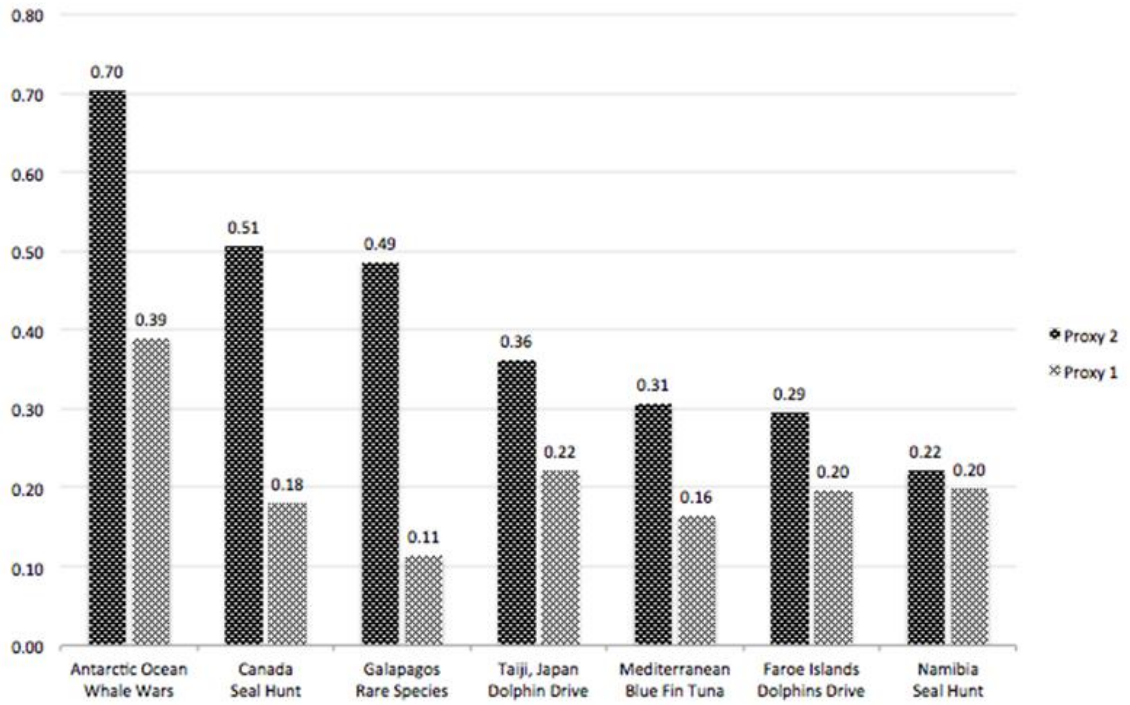
Trade-off parameters (weights) for both the AVE and the ARG 2 stakeholder are presented in Table 8. There moderate agreement, Pearson r = 0.59. While the AVE gave the highest weight to minimizing illegal kills, the ARG 2 stakeholder gave the highest weight to maximizing media attention.

Table 8. AVE and Stakeholder Weights for ARG 2

	Swing Weights	
Attribute	Expert	Stakeholder
Min illegal kills	100	65
Max economic impact to the adversary	95	65
Min adversary (i.e. poachers) revenue	95	65
Max new policies protecting wildlife	70	25
Max media attention	70	100
Max donations	70	25
Max community education	30	75
Max recruitment	20	50
Min fines	15	0
Min direct costs of mission	15	0
Min expected losses	15	0

Expected utilities were calculated for both the AVE and the ARG 2 stakeholder using the AVE's score matrix, but different sets of weights and utility functions provided by each. Figure 8 presents the expected utilities for both the AVE (Proxy 1) and the ARG 2 stakeholder (Proxy 2). There is moderate agreement, Pearson r = 0.58. Notably, both the AVE model and ARG 2 stakeholder model identified the Antarctic whale campaign as the most preferred alternative.

Figure 8. AVE (Proxy 1) and Stakeholder (Proxy 2) Expected Utilities for ARG 2



5. CONCLUSIONS

Moderate to strong convergent validity was demonstrated in both case studies presented. In the first case study, the ARG 1 stakeholder developed a different MAU model and identified somewhat different alternatives. Although the ARG 1 stakeholder model was simpler (fewer objectives), the models overlapped quite a bit and included much the same concerns. Likewise, the ARG 1 stakeholder used somewhat different terminology for the alternatives, but there was a great deal of overlap and substantial agreement in preference rankings of the alternatives.

In the second case study, the ARG 2 stakeholder adopted the AVE MAU model and alternatives, but provided very different weights on the eleven attributes and quite different indications of risk attitude on the eleven single attribute utility functions. Despite these discrepancies in the two models, there was perfect agreement regarding the most preferred alternative among the seven alternatives considered. There was little agreement, however, regarding the remaining six alternatives, due to differences in attribute weights and utility functions between the ARG 2 stakeholder and the AVE.

Interestingly, in both cases, the ARG stakeholders placed substantially greater weight on the objective of media attention than did the either of the AVEs, who developed independent models based on published writings and website information. It seems reasonable that politically active groups would not advertise a fundamental objective such as maximizing media attention. Instead, both ARGs' published writings and web sites focused on more fundamental objectives, central to the core mission of the group. One could question whether media attention should even be considered a fundamental objective, since it is almost surely a means to a greater end in almost all cases. We raised this issue with both ARG stakeholders, and both were adamant that media attention is an ends unto itself. The premise seems to be that media attention is the equivalent to effecting change in the hearts and minds of the public. We suspect that most politically active groups place a great deal of weight on media attention that cannot be discerned from published materials. Care should be taken when developing adversary MAU models to consider hidden objectives of the group, and to account for objectives that are stated but perhaps underweighted.

Acknowledgements

We acknowledge the assistance of Kacie Shelton and Jacob Jacobsen for their assistance in constructing the proxy adversary objectives hierarchies and multi-attribute utility models. This research was funded by the U. S. Department of Homeland Security (DHS) through the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) under the cooperative agreement number 2010-ST-061-RE0001. Any opinions, findings, conclusions, and recommendations in this document are those of the authors and do not necessarily reflect views of the U. S. DHS or CREATE.

References

[1] R. S. John and H. Rosoff, H.. "Modeling effects of counterterrorism initiatives for reducing adversary threats to transportation systems," Journal of Homeland Security, Proceedings of the 2011 DHS Science Conference – 5th Annual University Network Summit, focused on Catastrophes and Complex Systems: Transportation, Washington, D.C., March 30–April 1, (2011).

[2] G. L. Keeney and D. v. Winterfeldt. "Identifying and structuring the objectives of terrorists" Risk Analysis, 30(12), pp. 1803-1816, (2010).

[3] H. Rosoff and R. S. John. "Decision analysis by proxy for the rational terrorist. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009), Workshop on Quantitative Risk Analysis for Security Applications (QRASA), Pasadena, California, July 11-17, (2009).

[4] S. D. Bond, K. A. Carlson, and R. L. Keeney. "Generating objectives: Can decision makers articulate what they want?," Management Science, 54(1), pp. 56-70, (2008).

[5] S. D. Bond, K. A. Carlson, and R. L. Keeney. "Improving the generation of decision objectives," Decision Analysis, 7(3), pp. 238-255, (2010).