# Dynamic vs. static decision strategies in adversarial reasoning

David A. Pelta[1] Ronald R. Yager[2]
1. Models of Decision and Optimization Research Group
Department of Computer Science and A.I., University of Granada,
C/Periodista Daniel Saucedo s/n, 18071 Granada, Spain
2. Machine Intelligence Institute, Iona College
New Rochelle, NY 10801, USA
Email: dpelta@decsai.ugr.es, yager@panix.com

***Abstract***— Adversarial decision making is aimed at determining optimal decision strategies to deal with an adversarial and adaptive opponent. One defense against this adversary is to make decisions that are intended to confuse him, although our rewards can be diminished. It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what we want is to *force the presence of uncertainty* in order to confuse the adversary in situations of repeated conflicting encounters. Using simulations, the use of dynamic vs. static decision strategies is analyzed. The main conclusions are: a) the use of the proposed dynamic strategies has sense, b) the presence of an adversary may produce a decrease of, at least, $45\%$ with respect to the theoretical best payoff and c) the relation between this reduction and the way the uncertainty is forced should be further investigated.

***Keywords***— Adversarial reasoning, uncertain environment, decision strategies, simulation

## 1 Introduction

Adversarial decision is largely about understanding the minds and actions of one's opponent. It is relevant to a broad range of problems where the actors are actively and consciously contesting at least some of each others' objectives and actions [1]. The field is also known as decision making in the presence of adversaries or adversarial reasoning.

In its most basic form, adversarial decision making involves two participants, white and black, each of which chooses an action to respond to a given event without knowing the choice of the other. As a result of these choices, a payoff is assigned to the participants. When this scenario is repeated many times, i.e. situations of repeated conflicting encounters arise, then the situation becomes complex as the participants have the possibility to learn the others strategy. Examples of this type can be found in the military field, but also in problems of real-time strategy games, government vs government conflicts, economic adversarial domains, team sports (e.g., RoboCup), competitions (e.g., Poker), etc. [1]

Adversarial decision making is aimed at determining optimal strategies (for white) against an adversarial and adaptive opponent (black). One defense against this adversary is to make decisions that are intended to confuse him, although white's rewards can be diminished.

It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what white wants is to make its behaviour as uncertain or unpredictable as possible. In other words, white wants to *force the presence of uncertainty* in order to confuse the adversary while its payoff is as less affected as possible.

In previous work [2], we proposed a model to study the balance between the level of confusion induced and the payoff obtained and we concluded that one way to produce uncertainty is through decision strategies for white that contain certain amount of randomness. Here we focus on learning strategies that white can use as a means of optimizing his payoffs in situations of repeated conflicting encounters. Essentially we are studying how white can defend against an opponent who is trying to learn their decision rules.

In this paper, we want to analyze the case where white's decision strategy is not constant along the time, but modified following certain rules. We explore two alternatives, one is to vary the number of candidates alternatives in terms of their associated payoffs and second is based on the basic concept of $\alpha$-cuts, where the value of $\alpha$ is varied.

The contribution is organized as follows: some basic concepts on adversarial reasoning are outlined in Section 2. Then, Section 3 describes the main characteristics and components of the model used. Section 4 introduces static decision strategies for both agents and then, shows how they can be transformed into dynamic ones. In Section 5 we describe the computational experiments performed and the results obtained and finally, Section 6 is devoted to discussions and further work.

## 2 Adversarial Reasoning

As stated before, adversarial decision making is largely about understanding the minds and actions of one's opponent. A typical example is the threat of terrorism and other applications in Defense, but it is possible to envisage less dramatic applications in computer games where the user is the adversary and the computer characters are provided with adversarial reasoning features in order to enhance the quality, hardness and adaptivity of the game. The development of intelligent training systems is also an interesting field.

The threat of terrorism, and in particular the 9/11 event, fueled the investments and interest in the development of computational tools and techniques for adversarial reasoning. However, the field has earlier developments. For example, almost twenty years ago, P. Thagard [3] states

> In *adversarial problem solving*, one must anticipate, understand and counteract the actions of an opponent. Military strategy, business, and game playing all require an agent to construct a model of an

ISBN: 978-989-95079-6-8

opponent that includes the opponent's model of the agent.

Game theory is perceived as a natural good choice to deal with adversarial reasoning problems. For example, a brief survey of techniques where the combination of game theory with other approaches is highlighted, jointly with probabilistic risk analysis and stochastic games is presented in [4]

However, nowadays it is assumed that the field transcends the boundaries of game theory [1]. As stated in [5]: "we argue that practical adversarial reasoning calls for a broader range of disciplines: artificial intelligence planning, cognitive modeling, control theory, and machine learning in addition to game theory. An effective approach to problems of adversarial reasoning must combine contributions from disciplines that unfortunately rarely come together".

## 3 The model

The framework used to conduct our study is a slight modification of the previously proposed in [2].

It is based on two agents white $W$ and black $B$ (the adversary), a set of possible inputs or events $E = \{e_1, e_2, \ldots, e_n\}$ issued by a third agent $R$, and a fuzzy set of potential responses or actions $A_i = \{a_1, a_2, \ldots, a_m\}$ associated with every event. These fuzzy sets are organized as rows in a matrix $P$ as :

$$P(n \times m) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ p_{31} & p_{32} & \cdots & p_{3m} \\ \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix}$$

where $p_{ij} \in [0, 1]$ is the level of suitability of action $j$ to respond to the event $i$. More precisely, $p_{ij}$ is the degree of membership of action $j$ to the fuzzy set of *suitable actions* associated with event $e_i$. We do not require these fuzzy sets to be normalized.

Agent $W$ has a strategy to decide which action to take given a particular event $e_k$ and perfect knowledge of matrix $P$. The aim for $W$ is to maximize the sum of the profits or payoffs given a set of inputs. These inputs or events are issued one at a time by $R$ and, in principle, they are independent.

The payoff of a given action is proportional to its suitability with respect to the given event. For the sake of simplicity, in this contribution we assume that the payoff is the value of suitability.

Agent $B$ wishes to learn the actions that $W$ is going to take given a particular input $e_k$ so as to reduce agent $W$ payoff. Agent $B$ does not know matrix $P$. It has access to the decisions made previously by $W$ and, if the guess matched the decision taken by $W$, then $B$ obtains some reward. We may think the situation as an "imitation game", where the aim for $W$ is to avoid being imitated.

A graphical view of the model is shown in Figure 1 while the whole procedure is described in Algorithm 1.

The payoff's calculation for $W$ at stage $j$ is defined as:

$$p' = p_{jk} \times F(a_g, a_k) \tag{1}$$

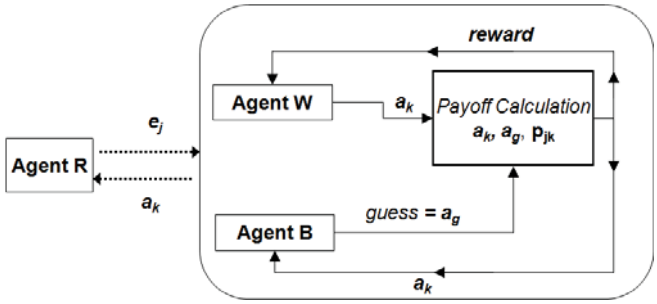



Figure 1: Graphical representation of the model. Events $e_j$ are issued by agent $R$ while response or actions $a_k$ are taken by agent $W$.

**Algorithm 1** Sequence of steps in the model.

**for** $j = 1$ to $N$ **do**
  A new event $e_j$ arises.
  Agent $B$ "guesses" an action $a_g$
  Agent $W$ determines an action $a_k$
  Calculate payoff for $W$
  Agent $B$ records the pair $e_j, a_k$
**end for**

where $F$ is:

$$F(a, b) = \begin{cases} 0 & \text{if a = b} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

As stated before, the aim for $W$ is to maximize the sum of the payoffs. In other words, a strategy for $W$ should be aimed at defining a sequence of actions that gives the higher payoff while avoiding being correctly guessed.

The terms *event* and *action* used here should be understood in a broad sense. For example, an event may represent a particular simulation scenario while an action may represent a full plan. Also, the actions may represent a Dempster-Schaffer belief structure to reflect the ideas posed in [6].

## 4 Modeling the Behavior of the Agents

In this section, we provide alternatives for modeling the behavior of both agents. For simplicity, we assume that the inputs issued by agent $R$ are equiprobable and that the number of inputs equals the number of actions (i.e, the payoff matrix is square).

### 4.1 Strategies for Agent $B$

Agent $B$ applies a very simple frequency-based decision strategy. We define a matrix of observations $O$ with $M \times M$ dimensions, where each $O_{ij}$ stores the number of times that action $i$ was observed (from $W$) when the event was $e_j$. Given an event $e_j$, the following decision strategy for $B$ is used:

*Proportional to the Frequency (PF)*: the probability of selecting an action $i$ is proportional to $O_{ij}$ (the observed frequency from agent $W$) [2].

### 4.2 Strategies for Agent $W$

Agent $W$ knows he is being observed, so the idea is to change its behavior in order to confuse the adversary. The agent needs to take sub-optimal decisions in order to get benefits in the long term.

ISBN: 978-989-95079-6-8

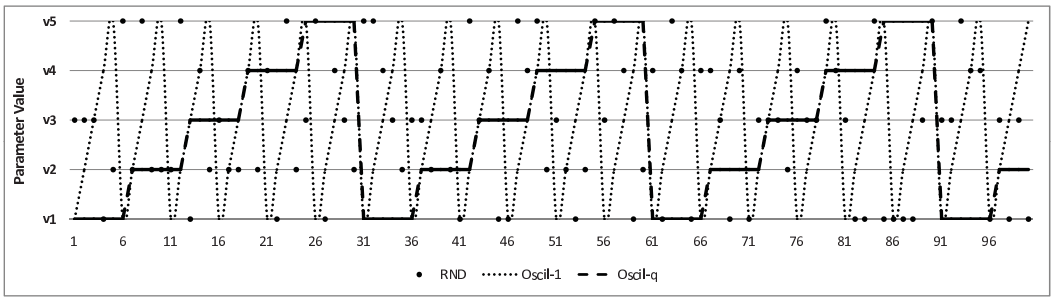



Figure 2: Adaptation schemes for the parameters in the dynamic strategies. The example shows the case where a parameter takes values from a set of possible discrete alternatives $\{v_1, v_2, v_3, v_4, v_5\}$

Given an observed input $e_i$ and the suitability matrix $P$, we propose two basic strategies:

1. *Random-Among $< k >$ Best Actions (R-k-B)*: Select randomly an alternative among the $k$ most suitable ones [2].
2. *Random-From $\alpha$-cut (R-$\alpha$)*: select randomly an alternative among those with a minimum level of suitability $\alpha$.

We remark here that these strategies are not *mixed strategies* in the sense of game theory as we do not define a probability distribution over the set of alternatives. Once a subset of alternatives is selected, then any of them can be selected with equal probability.

It is clear that both parameters, $k$ and $\alpha$, can be used to control the amount of uncertainty in the action selection. Some specific values for those parameters lead to interesting behaviours:

- $k = 1$: always select the most suitable action.
- $k = M$: any action is considered equally suitable.
- $0 < k < M$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Particular cases for $\alpha$:

- $\alpha = 1$: always select the action whose level of suitability is 1. This is not a good strategy as it is easily learnable.
- $\alpha = 0$: any action is considered equally suitable.
- $0 < \alpha < 1$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Although at first sight both strategies look similar, there is an important difference. The former strategy (*R-k-B*) is independent of the suitability scale: it just selects the $k$ most suitable alternatives and then takes a decision. The later strategy is based on $\alpha$-cuts, so if the value of $\alpha$ is high, then it may happen that the subset of alternatives of the corresponding $\alpha$-cut becomes empty. In this case, an alternative would be randomly selected.

#### *4.2.1 Dynamic Strategies*

When $\alpha$ and $k$ are assigned specific values, then a "static strategy" is obtained. Here, we propose a set of "dynamic strategies" where the values of the control parameter are varied along the time following the patterns shown in Fig. 2.

In order to adapt the parameter $k$, we propose three schemes:

1. *RND*: at each stage, $k$ is a value from a uniform distribution in $[1, M]$.
2. *Oscil-1*: after each event, $k = k + 1$. When $k = M + 1$ then $k = 1$.
3. *Oscil-10*: the same as before but the value of $k$ is changed after ten events.

For the adaptation of $\alpha$ we propose similar schemes:

1. *RND*: at each stage, $\alpha$ is selected randomly from the set $V = \{0.15, 0.3, 0.45, 0.6, 0.75\}$
2. *Oscil-1*: after each event, $i = i + 1$, $\alpha = V[i]$. When $i = M + 1$ then $i = 1$
3. *Oscil-10*: the same as before but the value of $\alpha$ is changed after ten events.

## 5 Experiments and Results

The aim of the experiment is to evaluate how $W$'s payoff is affected when dynamic vs. static decision strategies are used.

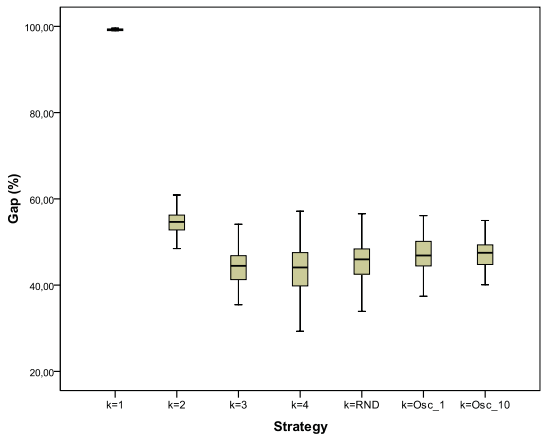
In order to do this, the following considerations are taken. We fix the number of events and alternatives to $M = 5$. Then, for every $W$'s strategy we made 100 repetitions of the scheme shown in Algorithm 1, where $N = 500$. The matrix $P$ is randomly generated each repetition. In this way we avoid potential biases due to particular configuration of values in $P$. At the end of each repetition, we record:

- $Gap$: the gap (as percentage) between the payoff obtained by $W$ and the optimum (calculated as the sum of the payoffs associated with the best action for every event),
- $Guess$: the number of times that $B$ correctly guessed the action of $W$ (also as percentage with respect to 500 actions).

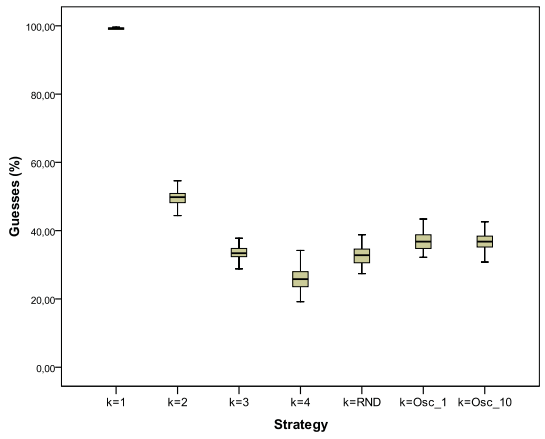
For both cases, the lower the values are, the better the performance of the strategy is.

We evaluated six dynamic strategies (three for $\alpha$ and three for $k$), four static strategies *R-k-B* using $k = 1, 2, 3, 4$ and five static strategies *R-$\alpha$* using $\alpha = \{0.15, 0.3, 0.45, 0.6, 0.75\}$.

ISBN: 978-989-95079-6-8




(a)




(b)

Figure 3: Strategy *R-k-B*: average $Gap$ (a) and $Guess$ (b) for each static and dynamic configuration

### 5.1 Results

In first place we will analyze the results for the *R-k-B*. Figure 3 shows boxplots for the average $Gap$ and $Guess$ (the average number of correct guesses) when the parameter $k$ is fixed or adapted during the repetition.

As it also occurred in [2], the value $k = 1$ is the worst alternative in terms of both measures. In this case, the best action is always selected, thus after a few events the frequency of those actions in the observations matrix kept by $B$ are the only ones different than zero and $B$ always choose them. As $k$ increases, the performance is better on average but the standard deviation is increased. The proposed adaptive schemes for $k$ do not show big differences among them. What is clear, is the reduction in the standard deviation from the $k = RND$ strategy to the one that force oscillations for $k$ every ten events. In terms of the $Guess$ measurement, it is interesting to note that a variation of this single parameter can decrease the number of correct predictions from $50\%$ when $k = 2$ to a lowest value of $25\%$ when $k = 4$.

In order to assess if the differences among the strategies in the average gap have statistical significance, we performed an ANOVA test followed by a post-hoc analysis using Tamhane's test with $p < 0.05$. The results are shown in Table 1; the signs indicate that the average gap between strategies $(i, j)$ (row,col) is different with statistical significance. A '+ sign indicates that strategy $i$ is better, while a '- denotes that $j$ is better. Absence of symbol indicates that the difference had no statistical significance. In this case, the best results are obtained by static strategies with parameters $k = 3$ and $k = 4$. The strategy $RND$ shows a similar performance, however it is not so good to outperform the other dynamic strategies.

| *R-k-B* | 1 | 2 | 3 | 4 | RND | Osc-1 | Osc-10 |
|---|---|---|---|---|---|---|---|
| 1 | | - | - | - | - | - | - |
| 2 | + | | - | - | - | - | - |
| 3 | + | + | | | | + | + |
| 4 | + | + | | | | + | + |
| RND | + | + | | | | | |
| Osc-1 | + | + | - | - | | | |
| Osc-10 | + | + | - | - | | | |

Table 1: Summary of the statistical testing for *R-k-B* strategies. See text for details.

| *R-*$\alpha$ | 0.15 | 0.30 | 0.45 | 0.60 | 0.75 | RND | Osc-1 | Osc-10 |
|---|---|---|---|---|---|---|---|---|
| 0.15 | | | + | + | + | | | |
| 0.30 | | | + | + | + | | | - |
| 0.45 | - | - | | + | + | - | - | - |
| 0.60 | - | - | - | | + | - | - | - |
| 0.75 | - | - | - | - | | - | - | - |
| RND | | | + | + | + | | | |
| Osc-1 | | | + | + | + | | | |
| Osc-10 | | + | + | + | + | | | |

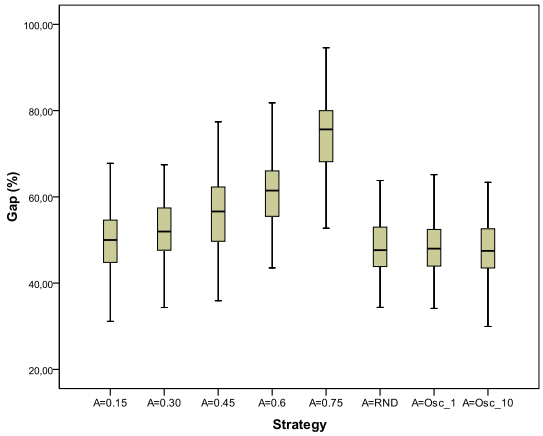
Table 2: Summary of the statistical testing for *R-*$\alpha$ strategies. See text for details.

A similar analysis can be done for the *R-*$\alpha$ strategy. Figure 4 shows boxplots for the average $Gap$ and $Guess$ when the parameter $\alpha$ is fixed or adapted during the repetition. The label $A$ in the plots stands for $\alpha$.

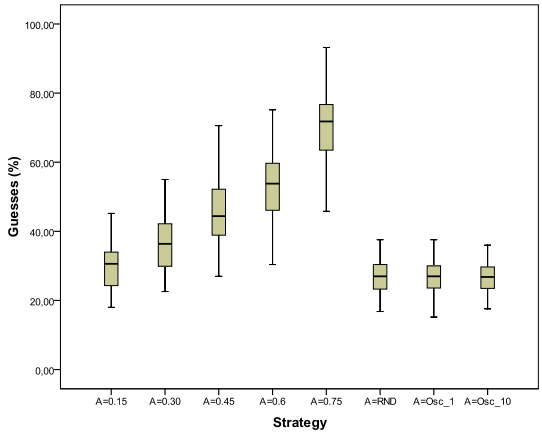
As $\alpha$ increases, the performance becomes worst. The reason is related with the way the matrix $P$ is generated. It may happen that no action has a level of suitability with a degree of membership higher than $\alpha$. In other words, the corresponding $\alpha$-cut leads to an empty set of alternatives. In this case, a pure random decision strategy is taken. The counterpart is that now, the proposed adaptive schemes for $\alpha$ are clearly better than the static alternatives. Between the adaptive schemes, no clear differences appeared.

As before, in order to assess if the differences among the strategies in the average $Gap$ have statistical significance, we performed statistical testing. The results are shown in Table 2. It can be confirmed that when using a static strategy, lower values of $\alpha$ like 0.15 or 0.3 led to better results than those obtained with higher ones. However, any of the dynamic strategies proposed, even $RND$, obtained the same average $Gap$. Moreover, the strategy *Osc-10* (that changed the $\alpha$ value every ten events) provided better performance than all the static strategies, excepting $\alpha = 0.15$.

ISBN: 978-989-95079-6-8




(a)




(b)

Figure 4: Strategy *R*-$\alpha$: average $Gap$ (a) and $Guess$ (b) for each static ($A$ stands for $\alpha$) and dynamic configuration

It is reasonable to assume that a relation exist between the number of correct guesses and the payoff obtained. If the former is high, it is clear that the payoff should be low (this is the case when $k = 1$). But what happen when the number of correct guesses is low? What are the payoff values that can be reached?. Figure 5 shows the average of each measure for every strategy (the case when $k = 1$ is omitted for visualization purposes). Within each kind of strategy (*R-k-B*, R-$\alpha$), the alternatives are ordered increasingly in terms of $Gap$.

It is clear that just reducing the number of guesses is not enough to improve the gap. For example, *A=Osc-10* achieved a lower value of $Guess$ than $k = 3$ but a higher value of $Gap$. This was also already notice in [2], where a purely random strategy led to the lowest value of $Guess$. Interestingly, all the cases where the curve for $Guess$ stands below the one for $Gap$ correspond to strategies based on $\alpha$-cuts. The reason is not clear and further investigations are needed.

To conclude the analysis, Figure 6 shows several scatter plots displaying the relation between both measures. Every point correspond to a particular repetition, thus 100 point per plot are displayed. Plots on the left correspond to *R*-$\alpha$ strategy while those on the right to *R-k-B*. Three static values are shown per each parameter, while the bottom plot corresponds to the best dynamic strategy. The X axis is the $Gap$ while Y axis shows the measure $Guess$.

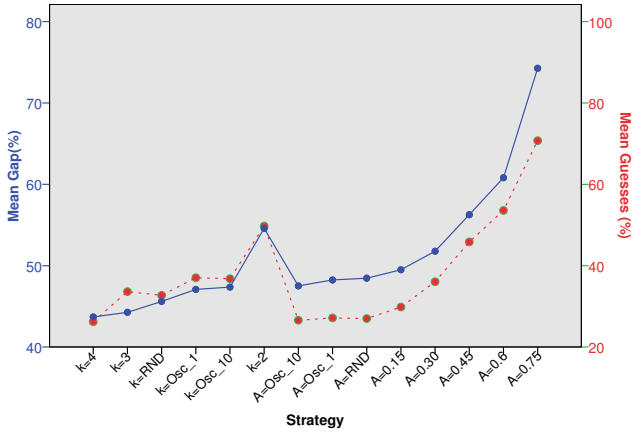



Figure 5: Average $Gap$ (left Y axis) and $Guess$ (right Y axis) for every strategy. The dotted line is used for $Guess$.

Several elements can be observed in the Figure. Let's start with the *R*-$\alpha$ strategy (left). As $\alpha$ increases, we can observe a progressive concentration of the points towards the upper right corner of the plot. This means that the results are getting worse: more guesses, higher gap (agent $W$ is getting much lower payoff than the potential optimum). It is noticeable the high variation the results are showing. For instance, when $\alpha = 0.45$ (second plot, in the left), one can observe simulation with guesses around $30\%$ and gap around $30\% - 45\%$ and others with guesses higher than $60\%$ and gap higher than $60\%$. When a simple dynamic strategy is used ($\alpha = RND$), then the number of correct guesses is almost always below $30\%$. However, the range of values that can be obtained for gap value is quite wide, going from $30\%$ to $60\%$.

The plots are a bit different for the *R-k-B* strategy. As $k$ increases, the percentage of guesses decrease. This is reasonable taking into account how the strategy works: take the $k$ best alternatives and then, choose random. In the experiments performed, the total number of actions is 5, so, when $k = 4$, then just the worst strategy is eliminated. In other words, as $k$ increases, agent $B$ may conclude that $W$ is behaving randomly. In turn, the gap value is more variable, ranging from $35\%$ to $55\%$ approx.

## 6 Discussion and Future Work

In this work we focused in the context of adversarial reasoning where a player $W$ wants to *force the presence of uncertainty* in order to confuse the adversary $B$ while its payoff is as less affected as possible in situations of repeated conflicting encounters. We extended a previous work where the decision strategies were fixed along the time, to consider dynamic strategies: the way the decisions are taken varies with the time.

When the decision strategy is based on $\alpha$-cuts, then a dynamic variation of $\alpha$ led to better results than the static counterpart. Moreover, none of the adaptation schemes proposed led to worst results than those obtained by a specific configuration of the control parameter. However, in the $R - k - B$ case, the results are not so clear but still good. The best results are obtained with $k = 3, k = 4$. Using any of these values, the strategy obtained performed better than four of the other

ISBN: 978-989-95079-6-8

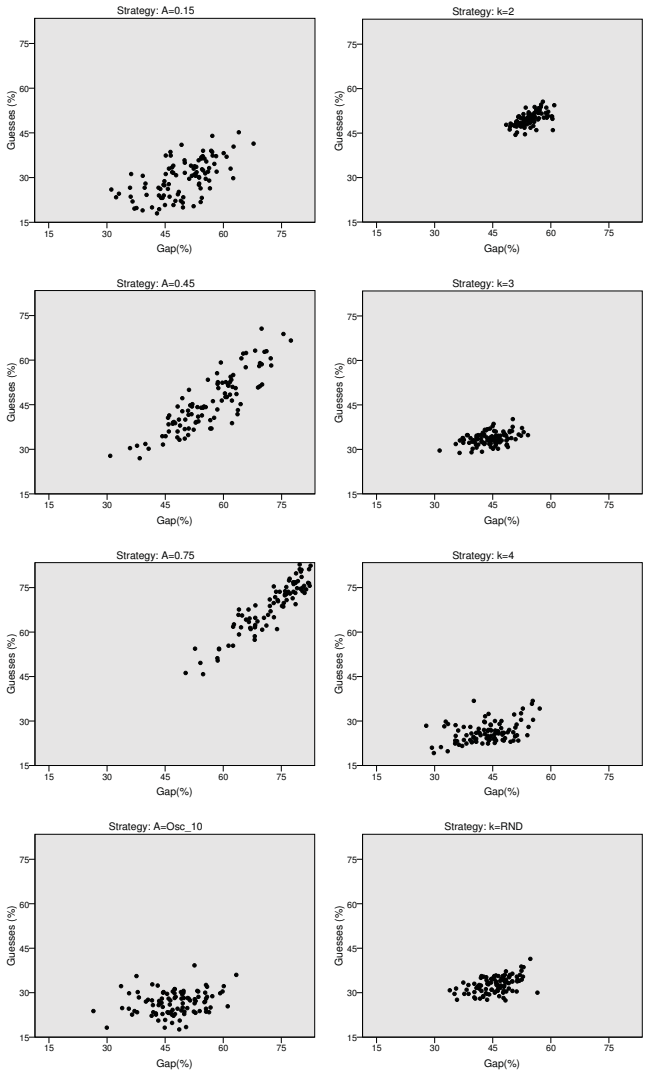



Figure 6: Every point represents the result $(Gap, Guess)$ of a repetition.

strategies available. The best dynamic alternative is $RND$ which assigns a random value to $k$ before applying the strategy. This strategy achieved a similar performance (on average) with respect to $k = 3, k = 4$.

The use of dynamic strategies, besides its potential performance, has another benefit which is to avoid determining a specific value for the control parameter. This is not a trivial matter as there is no simple way to infer which setting will provide the best value for a different simulation scenario.

The study performed also revealed an interesting point: the lowest average $Gap$ value for a dynamic strategy was $45\%$ when using $k = RND$ and this implies that $W$'s payoff is $45\%$ lower than the theoretical optimum one. In [2], we showed that using a purely random decision strategy, the gap is higher than $50\%$. The question now is: would it be possible to design a strategy with guaranteed (in a formal sense) performance?.

The way the dynamism is included in this proposal is quite simple, but several alternatives are available for producing improvements. Now, agent $W$ is not analyzing the payoff obtained against the one expected (which is stored in matrix $P$). Agent $W$ can use the fact that its payoff is being affected in order to produce an intelligent adaptation of its strategy. In this context, the use of fuzzy rules may play an important role as it would help to model adaptation strategies as: *if the reduction of payoff is high, then increase uncertainty*, or *if this event occurred a high number of times then increase uncertainty*. Other lines of research are related with the model used. Here, the way the payoff matrix is generated may affect the results obtained by the strategy based on $\alpha$-cuts as the differences between the suitability of alternatives may be low. Other ways to define such matrix may be necessary. Also, the sequence of events is completely random now, but other options are available.

The relation between the number of correct guesses and the gap value is clear in some situations but in others (as shown in Fig. 6), it is possible to obtain a broad range of gap values for the same number of guesses. In this context, it seems not trivial to use the number of guesses as a predictor for the gap value. A different alternative may take into account that the presence of uncertainty is reflected in the matrix of observations $O$ that agent $B$ construct, so it would be interesting to look for correlations between the gap value and some measures about this matrix $O$. The ideas posed here are left as future work.

Finally, we would like to mention that one of the reviewers claimed that the problem posed here can be solved with the standard tools of game theory arguing that the problem can be seen as a sequence of $n$ independent non cooperative games. However, we do not think that this is the case as, for example, the matrix of observations changes at each step, we are not dealing with mixed strategies and so on. Moreover, we claim that this kind of analysis can be applied when the events are correlated in some way, and also, when more sophisticated adaptation and learning mechanisms are considered in both agents.

## Acknowledgments

This work is supported in part by projects TIN2008-01948 from the Spanish Ministry of Science and Innovation, and P07-TIC-02970 from the Andalusian Government.

### References

[1] A. Kott and W. M. McEneany. *Adversarial Reasoning: Computational Approaches to Reading the Opponents Mind*. Chapman and Hall/ CRC Boca Raton, 2007.

[2] D. Pelta and R. Yager. On the conflict between inducing confusion and attaining payoff in adversarial decision making. *Information Science*, 179:33–40, 2009.

[3] P. Thagard. Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science*, 16(1):123 – 149, 1992.

[4] E. Kardes and R. Hall. Survey of literature on strategic decision making in the presence of adversaries. Report 05-006, National Center for Risk and Economic Analysis of Terrorism Events, 2005.

[5] A. Kott and M. Ownby. Tools for real-time anticipation of enemy actions in tactical ground operations. In *Proceedings of the 10th International Command and Control Research and Technology Symposium*, 2005.

[6] R. R. Yager. A knowledge-based approach to adversarial decision making. *International Journal of Intelligent Systems*, 23:1–21, 2008.

ISBN: 978-989-95079-6-8