On the learning patterns and adaptive behavior of terrorist organizations

On the Learning Patterns and Adaptive Behavior of Terrorist Organizations*

Johannes G. Jaspersen†, Gilberto Montibeller‡

September 4, 2019

Abstract

The threat to national security posed by terrorists makes the design of evidence-based counter-terrorism strategies paramount. As terrorist organizations are purposeful entities, it is crucial to understand their decision processes if we want to plan defenses and counter-measures. In particular, there is evidence that terrorists are both adaptive in their behavior and driven by multiple objectives in their actions. In this paper, we use insights from learning theory and compare several different reinforcement learning models regarding their ability to predict terrorists' actions. Using data on target choices of terrorist attacks and two different objectives (renown and revenge), we show that a total reinforcement learning with power (Luce) choice probabilities and information discounting can be used to model the adaptive behavior of terrorists. The model renders out-of-sample predictions which are comparable in their validity to those observed for learning in laboratory studies. We draw implications for counter-terrorism strategies by comparing the predictive validity of the different models and their calibrated parameters. Our results also offer a starting point for studying the convergence process in game theoretic analyses of conflicts involving terrorists.

Keywords: decision processes; behavioral OR; OR in defense; adversarial risk analysis, decision analysis.

*We are thankful for the comments from the participants at the 2016 European Conference on Operational Research, the 2016 Society for Risk Analysis Annual Meeting, the 2017 INFORMS annual meeting and the 2017 Munich Workshop on Behavior and Terrorism as well as the seminar participants at the Universities of Tübingen, Hamburg and Munich. We are indebted to Richard Dyson (the editor of the Journal), three anonymous reviewers, Claude Berrebi, Richard John, Edward Kaplan, Stefan Neuß, Richard Peter, Sumedh Shastri and Justin Sydnor for helpful comments. Tara Kuruvila, Joëlle Näger and Antonio Orsulic provided valuable research assistance. The usual disclaimer applies.

†Munich Risk and Insurance Center (MRIC), LMU Munich, E-Mail: jaspersen@bwl.lmu.de, Phone: +49 89 2180 2792.

‡Corresponding author. School of Business and Economics, Loughborough University, E-Mail: g.montibeller@lboro.ac.uk, Phone: +44 1509 223267.

1 Introduction

Given the permanent threat to national security created by foreign and domestic terrorists, the design of evidence-based counter-terrorism strategies is becoming increasingly important in intelligence operations research (Kaplan, 2012; Keller and Katsikopoulos, 2016; Seidl et al., 2016). As in all other operations research applications involving the modeling of agents (Franco and Hämäläinen, 2016), assumptions about the underlying behavior of terrorist organizations are necessary for devising such strategies (Bhashyam and Montibeller, 2015). The idea of the maniac terrorist as an irrational actor has mostly been rejected in the literature (Caplan, 2006). Instead, terrorists are seen as purposeful agents with multiple objectives (Richardson, 2006) and engaged in adaptive decision processes (Arin et al., 2011). The multi-objective nature of their action has led to suggestions of modeling their objectives using multi-attribute value functions (Bhashyam and Montibeller, 2012; Siebert et al., 2016). How exactly their decision processes should be modeled is, however, unclear. Terrorist organizations often have a large strategy space (Konrad, 2004) and do not possess full information about the situation they operate in (Powell, 2007). Adaptive processes based on reinforcement learning (Camerer and Ho, 1999; Gill and Prowse, 2016) thus seem a likely candidate to model their behavior.[1] However, as far as we are aware of, the nature of such adaptive processes has not been explored in this decision context.

In this paper, we investigate which model of learning and behavioral adaptation best describes the multiple objectives decision process of terrorist organizations. Specifically, we study terrorists' target choices for bombings. Using evidence from global terrorism incidents and data from Google Trends, we analyze which of the models commonly proposed for reinforcement learning and behavioral adaptation in repeated interaction describes behavior best. We derive insights about the adaptive behavior of terrorist organizations by describing the behavioral implications of each tested model. Moreover, future studies aiming at deriving counter-terrorism strategies can use our results to make assumption on the behavioral model of terrorist organizations. For this purpose, we offer a starting point both in terms of the chosen learning model and in terms of its parameterization.

Our results show, that among the tested models, a learning process based on total reinforcement leads to the best description of the adaptive behavior of the analyzed organizations. That is, terrorist organizations seem to react to the total sum of payoffs gained from previous trials. We show that the model renders a fit similar to the one observed in behavioral laboratory studies of repeated games. Our findings provide evidence for the law of effect (Thorndike, 1927) and for the power law of practice (Newell and Rosenbloom, 1981) in terrorist organizations' learning processes. Additionally, we find that terrorist organizations discount prior experience at the expense of more recent reinforcements. Contrary to common models of aspiration based learning (e.g., Jaspersen and Peter, 2017), we find no evidence that an aspiration level or reference point drives the behavior of terrorist organizations.

[1] The terminology used here will describe learning as the cognitive process of forming estimates over the value of an action and adaptation as the behavioral consequence of this learning.

Our study makes three contributions to the literature. Firstly, we offer insight into the processes that govern the decisions of terrorist organizations. By identifying factors such as information discounting or the power law of practice, we increase our knowledge on how such groups decide. Even if future models of conflicts do not take the perspective of the terrorist organization as an adaptive entity learning from reinforcement, researchers can still integrate these partial aspects in their model. Showing the existence of behavioral motives themselves also offers some general insights into how effective counter-terrorism strategies can be designed, which we discuss in the last section of the paper.

The second contribution of our analysis is to the game theoretic literature on conflicts. While studying game theoretic equilibria is common for the analysis of conflicts involving terrorism (see, e.g., Zhuang and Bier, 2007; Pita et al., 2008; Alpern et al., 2011), it is often unclear how these equilibria are reached. Our study provides empirical evidence on the assumptions necessary for modeling the game theoretic convergence phase in counter-terrorist - terrorist interactions. Even though the final equilibrium that is reached is similar for most suggested adaptation processes, the behavior before the equilibrium can differ substantially (Camerer and Ho, 1999). For this reason, the choice of the learning process and behavioral adaptation assumed in the analysis of this convergence phase is crucial for the resulting prescriptive advice. In essence, this study provides the necessary tools for future modeling efforts which aim to get a more complete picture of the efficacy of counter-terrorism strategies. Importantly, such future studies will not be antithetic to analyses of the equilibrium of a game, but will rather complement them.

Prior OR studies of convergence processes in armed conflicts are commonly based on dynamic models (see, e.g., Kress and Szechtman, 2009; Kaplan, 2010; Seidl et al., 2016). These studies highlight the dynamic nature of the interaction between the different parties in a conflict and call attention to important features, such as the starting position of each party. The convergence process in these applications usually is the consequence of the dynamic model and is not due to a behavioral assumption itself. Our results can help future studies in this literature to incorporate behavioral decision theory into dynamic models. For instance, if one wanted to study a terror queue with multiple stock variables corresponding to different targets, the reinforcement learning model calibrated here could be used as the model which determines how terrorist organizations make the choice between the different targets.

Three other studies have empirically estimated models for describing the behavior of terrorist organizations. Bohorquez et al. (2009) and Johnson et al. (2011) analyze the timing and severity of terrorist attacks. They focus on modeling a conflict rather than an organization. These studies also focus mostly on large insurgent conflicts such as the current situations in Iraq and Afghanistan and are thus of limited value for informing western public policy. Arin et al. (2011) exclusively study the western context and calibrate a Markov chain model on terrorist activity and counter-terrorism spending in the UK. They show terrorist activity to be less likely to occur after military spending by the government has been increased. This implies both reactive behavior of terrorist organizations and some benefit to counter-terrorism spending. Their model is, however, only of limited help

for understanding adaptation in game theoretic settings, since Markov chains are by definition independent of past events. Therefore, this paper is, as far as we are aware, the first to relate experiential learning models to the behavior of terrorist organizations. We do not, however, claim that these models necessarily represent the psychological processes that terrorists follow. Instead, we hope that learning models can help in understanding and predicting terrorist organizations' adaptive behavior (see Katsikopoulos, 2014, for a discussion on the role of psychological processes in descriptive models).

The paper proceeds as follows. In the next section, we introduce our specific research setting and the data and empirical strategy used for the analysis. The third section presents the different models that were analyzed as well as their calibrations. In the fourth section, we discuss the robustness and limitations of our analysis. The last section discusses the findings, draws implications for counter-terrorism strategies and concludes.

2 Research Setting and Data

As mentioned in the introduction, we investigate which model of reinforcement learning and behavioral adaptation best describes how terrorist[2] organizations select targets for bomb attacks. We focus on bomb attacks, because they have been the most numerous in the recent decades. Of the 142,714 terrorist attacks between 1968 and 2014 documented in the Global Terrorism Database (GTD, START, 2015), almost half (68,590 or 48.06%) had a bombing as its primary attack type and another 5,061 attacks featured at least some kind of explosive device. However, our methodology can be applied to any kind of attack and while it is unclear whether organizations learn across different attack types, there is little reason to suspect that terrorist organizations adapt their behavior differently for different kinds of attacks. The learning models tested here were developed for situations which are highly similar across trials. By focusing solely on bomb attacks, we attempt to approximate a setting of repeated homogeneous decision situations with our data as best as possible. To check whether our results hold for a broader scope with less homogeneity between attacks, we also conducted a robustness analysis using both bomb attacks and armed assaults.

When empirically analyzing target selection, it is common to group certain types of targets together to form broader categories (Santifort et al., 2013). The GTD lists 22 different target types. We group these into five categories: Business, Government, Military, Private Citizens and Infrastructure. As is common in the literature, we removed all attacks from the analysis that did not have a known target type or that only targeted other insurgent groups. The latter exclusion is made because conflicts between different insurgents are usually not the focus of counter-terrorism policy. Attacks with at least one non-excluded target type were included in the analysis. The

[2]We use the term *terrorist* without any implied value judgment. In fact, some of the organizations included in our data have later emerged as legitimate political factions in their countries of operation. We adopt the definition of terrorists from the Global Terrorism Database (START, 2015) which defines them as non-state actors using violence or the threat of violence to reach an ulterior goal which is not strictly monetary and targets a demography larger than the immediate group of victims of their actions.

complete categorization is given in Table 1. While most choices we made are straightforward, our categorization of religious figures and institutions as private citizens might be arguable. Since we did not see them as either infrastructure or businesses, private citizens seemed the most logical categorization. Abortion clinics are also most often attacked due to religious reasons. We thus included them in the same category as religious figures and institutions.

Table 1 – Categorization of Target Types

Type in Analysis	Type in GTD
Business	Business (*94.96%*); journalists & media (*4.32%*); tourists (*1.01%*)
Government	Government (general) (*91.54%*); government (diplomatic) (*9.13%*)
Military	Military (*53.31%*); police (*48.85%*)
Private Citizens	Private citizens & property (*94.96%*); abortion related (*0.00%*); religious figures/ institutions (*7.17%*)
Infrastructure	Airports & aircraft (*3.87%*); educational institution (*21.11%*); food or water supply (*1.86%*); maritime (incl. ports and maritime facilities) (*0.85%*); NGO (*2.40%*); telecommunications (*7.35%*); transportation (other than aviation) (*31.48%*); utilities (*31.40%*)

The table displays how target types in the GTD are aggregated to target types in our analysis. For each analysis target type, we list the shares of the GTD target types that make up each category. These shares differ between the different samples used in the analysis. The shares displayed here are those from the sample with both fatalities and public attention payoffs which is the sample used in the behavioral motives analysis (Section 3.2) and the predictions across groups analysis (Section 4). Four target types from the GTD were excluded in our analysis. These were the types "other", "unknown", "terrorists/ non-state militias" and "violent political parties".

In our analysis, the decision making entity is the terrorist group. For the remainder of this analysis, we treat the terrorist organization as one homogeneous entity which makes decisions based on its entire past experience. While such an approach is common in organizational research (e.g., Herriott et al., 1985; Jaspersen and Peter, 2017), we are aware that organizations in general and terrorist organizations in particular are often not homogeneous (Neumann, 2016). In fact, it can even be rational for the planners of a terrorist organization to implement a heterogeneous decision making process which may ensure higher resilience towards counter-terrorism (Enders and Jindapon, 2010). Nevertheless, we maintain this assumption here both for the sake of simplicity in the analysis and to make results comparable between different organizations. Moreover, the idea that every decision making cell in a given terrorist organization learns from all actions of that organization is not unrealistic, even if the group is organized as a highly fragmented network. Since all attacks listed in the GTD are sourced from public news reports about them, every decision making cell within the organization has the ability to learn about prior attacks and their outcomes even if it was not involved in the planning process.

In order to ensure the calibration of a coherent decision making process, we focused our analysis on those attacks which were carried out by identifiable and coherent terrorist organizations. We thus excluded attacks in which the perpetrators were only identified by an ideology or some other vague term such as "death squad". In total this led to the exclusion of 38,479 attacks by 14 different

perpetrators (additional information on the exclusions can be found in the Online Appendix). We also removed all organizations for which we only had data on less than 50 bombings. This was done to ensure that both short-term and long-term effects of the reinforcement learning process would be captured.

The core behavioral assumption in our analysis is that target choices of terrorist organizations are stochastic. These organizations choose a target for a bombing according to a distribution of choice probabilities and gain some payoff from the attack. A learning process accumulates this payoff into a reinforcement for the target type which in turn changes the probability distribution in the stochastic choice process for the next attack. The process is repeated several times. The decision for attacking one target or another is thus based on the past experience of the organization, specifically on the payoffs which were associated with attacks on such targets in the past. We thus model the decision making process as one under a high degree of causal ambiguity in the sense that actors are not weighting all possible consequences of their actions when making a decision, but instead act heuristically.

The reliance of our model on the reinforcements of actions leads to the question of which payoffs terrorist groups gain from their attacks. Terrorist organization often aspire rather vague, long-term objectives toward which the contribution of a single attack cannot sensibly be measured (Hoffman, 2006). However, these organizations also pursue a range of short-term objectives. Richardson (2006) lists three themes which together subsume these short-term objectives of terrorists: revenge, renown and reaction.

Revenge is the terrorists' desire to retaliate a real or perceived injustice to their constituency. Such injustice is most often represented by murder, be it direct (i.e., through violent action) or indirect (e.g., through support of a violent party or the establishment of an oppressive regime). As such, terrorist organizations often exert revenge through killing members of the opposing party. We thus use the number of killed victims in each attack to measure the payoff gained by the terrorist organization in terms of revenge.

Renown measures the fame and public attention gained for the terrorist organization. Renown can inspire public sympathy for the cause of the organization or have more mundane ends such as increased funding (Atran, 2003). While it could be argued that it is not the foremost goal of all terrorist organizations to kill as many people as possible, renown seems to be important for all of these organizations. Terrorism is the use of physical and psychological violence to achieve political goals (Hoffman, 2006). Such goals will most likely not be achieved if the populace is not aware of the organization and thus the reason the violence was committed.

We proxy the renown payoff by the relative number of Google searches made for the organization in the week after the attack. In 2010, Google was the leading search engine in all 37 countries in which organizations considered in our analysis committed bomb attacks and had an average market share of 91.4% (Kennedy and Hauksson, 2012). Google searches thus accurately approximate the interest of these countries' populations with access to the Internet. We assume that they also approximate the interests of the total populations, since we assume the political interests between

Internet users and non-Internet users to be at least positively correlated if not equal.

Reaction, lastly, is the terrorists' aspiration to achieve opponent behavior beneficial to their own cause. This can range from a simple ransom payment, to troop withdrawal or even the implementation of draconian countermeasures. The latter is sometimes a goal to reveal the "true face" of the opponent, which justifies the own cause and facilitates recruitment. Identifiable reactions that can be tied to a singular event are, however, extremely sparse. Additionally, they vary widely in their nature and it is not always clear whether the reaction elicited by an attack was the one actually intended. As such, we refrain from using this dimension of terrorists' objectives in our study. We are, however, aware that analyses of singular organizations or events might be enriched by including it in future studies.

Data on the terrorist attacks, including the number of fatalities and the perpetrator, were taken from the GTD (START, 2015). In addition to the aforementioned exclusions, we excluded those attacks which had missing data for the number of fatalities. 54 events in the GTD relevant to our analysis only had information on the year and the month, but not on the day of the attack. This was either due to circumstances in the investigation of the event (for instance, if a bomb was found on a certain day but it was unclear whether it was planted that day or some day before) or due to lacking data in the GTD. Such events were listed to have happened on the 0^{th} day of the month. We were able to acquire additional data or to approximate the date from the description in the GTD for 34 of those 54 events. For all events for which we could only find a timespan but no single day of attack, we used the midpoint of the timespan. For the other 20 events, we used the same logic and set the day of the month to the 15^{th} in order to bias the analysis as little as possible. All changes made to the data are listed in detail in the Online Appendix.

Data on Google searches were downloaded from Google Trends (`https://www.google.com/trends/`) on March 30^{th}-31^{st} 2016. They are independently and relatively scaled for each organization by indexing the search queries for the organization against other common search terms on a country-by-country basis. Weeks on Google Trends begin on Sunday. To make sure that we measure the number of searches for an organization after an attack, we used a linear interpolation of the searches in the week of the attack and the week following it. The weight attached to the first week was the number of days left in the week on the day of the attack. The weight of the second week was seven minus the weight of the first week. As Google Trends data is only available since 2004, our dataset is decreased in size if renown is used as a measure of reinforcement.

The total dataset is comprised of 20,170 bomb attacks by 69 organizations if only the fatalities are taken as payoffs. If only public attention proxy is taken into account, the data are reduced to 8,147 attacks by 28 organizations. If both payoff measures are combined, the number of attacks is further reduced to 7,928 because 219 attacks in the public attention analysis did not have information about the fatalities. The summary statistics for these three datasets are given in Table 2. It is interesting to notice that while both payoffs are positively correlated, this correlation is not particularly high at 7.24%. The two dimensions of reinforcement are thus related but do not measure the same thing. The learning periods describe the instances in which the model can update reinforcements

and thus ultimately choice probabilities through the mechanisms described in the next section.

Table 2 – Summary Statistics

	Nature of Payoff		
	Fatalities	Public Attention	Both
Organizations	69	28	28
Learning Periods	12,252	5,232	5,126
Incidents	20,170	8,147	7,928
Business[a]	*3,511 (17.41%)*	*705 (8.65%)*	*695 (8.77%)*
Government	*2,388 (11.84%)*	*909 (11.16%)*	*898 (11.33%)*
Military	*5,937 (29.43%)*	*3,269 (40.13%)*	*3,138 (39.58%)*
Private Citizens	*4,651 (23.06%)*	*2,926 (35.92%)*	*2,858 (36.05%)*
Infrastructure	*4,862 (24.11%)*	*1,308 (16.05%)*	*1,293 (16.31%)*
Average payoff	2.18	n.a.[b]	n.a.[c]
Standard deviation payoff	6.27[d]	9.49	n.a.
Payoff correlation		0.0724[e]	

The table depicts the summary statistics of the three sets of data used. Statistics in the left column describe the data in which the number of fatalities is taken as the payoff of an attack. Statistics in the middle column describe the data in which the relative number of Google searches is taken as the payoff of an attack. Statistics in the right column describe the data with both payoffs.
[a] Multiple target types per incident are possible.
[b] An average is meaningless since the variable is relatively scaled for each organization.
[c] The payoff structure is estimated in the analysis. Thus no descriptive statistics can be given.
[d] Describes the average standard deviation within each group.
[e] Reported for 27 of the 28 groups as the Corsican National Liberation Front never had any fatalities associated with a bombing.

3 Analysis

We are analyzing behavior of terrorist groups as the result of a stochastic learning process. We observe an agent's behavior over time with the current period being denoted as $t \in \{1, ..., T\}$. For the remainder of the paper, t counts only those days in which an attack occurred. In each period, each decision alternative $i \in \{1, ..., I\}$ has a probability $p_{i,t}$ ($\sum_{i=1}^{I} p_{i,t} = 1$) of being chosen. The literature on models of stochastic learning and behavioral adaptation can broadly be categorized into two different approaches: reinforcement learning and belief-based models. In this study, we use reinforcement learning, in which the estimated payoff distributions of actions depend only on the outcomes of the actions that the agents have experienced in past trials. In contrast to this are belief-based models, in which the agents use past experience to form beliefs about the future actions of their counteragents and then base their own choices on these beliefs.

Even though the literature is often seen as divided, it is sensible to assume that both prior reinforcements and beliefs about the counteragents play a role in the choices of terrorist organizations (see Camerer and Ho, 1999, for evidence of this in the neutral context of laboratory experiments.) We refrain from the combined approach for two reasons. Firstly, data on the actions of counterterrorist agencies is often hard or impossible to obtain. Secondly, even if this data were available

to us now, it is unclear how well the terrorists were able to observe the counterterrorist actions at the time they occurred. Applying belief-based models in this area would thus likely require modifications to the previously proposed models, which generally assume observable action by all agents. While we think this would be an extremely promising avenue of future research, we focus here on the less informationally demanding reinforcement learning models to provide a starting point for research on learning and behavioral adaptation of terrorist organizations.

When applying the reinforcement learning model, the choice probability is based on the reinforcement for this particular action at the time, $Q_{i,t}$ and the reinforcements of all other actions. The assumed learning process underlying this modeling assumption is that agents use their experience to gradually uncover the otherwise unknown outcome distribution of their choice alternatives (Hertwig et al., 2004). The reinforcements thus represent the limited information available to the agents about how good an action is. Behavioral adaptation then dictates that the better an action is perceived, the more likely the agent is to take it in future attacks. Several models for implementing reinforcement learning and behavioral adaptation exist. In the first instance, we will introduce the most common methods and show which one best describes the data. We then use this best fitting model in a second step, to further study the mechanisms involved in the reinforcement learning process of terrorist organizations.

3.1 Model Selection

We start the model selection analysis by introducing the different reinforcement models and the choice probability generation methods. We then introduce the empirical strategy and the results of the analysis.

3.1.1 Reinforcement Model

Reinforcement learning can work in two different ways. In the approach used here, the reinforcement of an action at time t, together with those reinforcements of all other actions, directly determines the choice probability of this action in period t. The alternative would be that the choice probability at time t is reached through updating the choice probability from the previous period, depending on the performance of last period's action. While this latter approach is not uncommon in organizational research (Jaspersen and Peter, 2017), we refrain from analyzing it here because it requires additional assumptions (see, e.g., Börgers and Sarin, 2000). We do, however, estimate such a model as a robustness check in Section 3.2.3.

Among the direct reinforcement models there are two predominant ways of modeling reinforcement accumulation. Roth and Erev (1995) propose reinforcements to simply be the sum of all previous payoffs, a model we will call *total* reinforcement model and that has received some empirical support in game theoretic applications (Erev and Roth, 1998). The alternative procedure is to consider *averaged* payoffs gained from action i (Mookherjee and Sopher, 1997). The difference in the underlying learning process is that in the total reinforcement model the agent's estimate of an action's outcome distribution becomes better with every time the action renders a positive payoff.

In contrast, in the average reinforcement model the sum of the payoffs gained from an action up that that point is considered in proportion to the number of trials the agent has experienced for that action. It is thus apparent that the latter model is more sophisticated than the former.

To model these two accumulation processes, we need to consider the idiosyncrasies of the analyzed empirical setting. While in laboratory tests of learning theory, each decision can be observed in an ordered and controlled way, the data from terrorist attacks is not so readily available. First we need to consider the nature of the time period t. We model it on a daily basis. Our models of reinforcement learning, however, do not model the timing of attacks (though models for this do exist, see Bohorquez et al., 2009; Johnson et al., 2011). We only consider the choice of bombing targets, conditional on the occurrence of an attack. The time period t thus counts those days in which there is an attack. We also think it is unrealistic to assume that terrorist organizations learn between attacks which are carried out on the same day. As such, reinforcement updating in our model takes place after each learning period, that is after each day with at least one attack. Since terrorist groups sometimes carry out more than one attack on a single day, the learning process can involve the reinforcements gained from multiple actions. Lastly, some attacks have multiple target types. Of the 20,170 bomb attacks in the data, 5.21% have two and 0.32% have three target types (or actions) associated with them. We take this into account when formulating the reinforcement learning model.

We assume the initial reinforcement of all actions to be equal to q. The total number of attacks in period t is denoted n_t and each attack has up to three actions associated with it which are collected in the set $A_{j,t}$ with $j \in \{1, ..., n_t\}$.[3] Let $\mathbf{I}_{con}$ be an indicator function with condition *con*, $|A_{j,\tau}|$ be the cardinality of set $A_{j,\tau}$ and $y_{j,t}$ be the payoff of attack j of time period t, then the total reinforcement model can be described as

$$Q_{i,t}^{total} = q + \sum_{\tau=1}^{t-1} \sum_{j=1}^{n_\tau} \mathbf{I}_{i \in A_{j,\tau}} \frac{y_{j,\tau}}{|A_{j,\tau}|}. \tag{1}$$

Note that we make the choice to distribute the payoff (revenge, renown, or both) of an attack between the different targets associated with it. Alternatively, we could attribute the full payoff to each action. That, however, would attribute more weight to attacks with multiple target types for which we see no empirical basis.

For the average reinforcement model, we make a similar assumption. An attack with multiple target types contributes to the averaged payoff of each target type only with the weight of $\frac{1}{|A_{j,t}|}$. The total number of attacks with target type i up until and including period t is thus calculated as

[3]The exact definition of n_τ warrants some discussion. Terrorist organizations sometimes attack multiple targets simultaneously such as planting multiple bombs at different government buildings at the same time. It is a matter of opinion whether such an action constitutes one ore multiple attacks. Since we want to minimize our own subjective judgments in the data preparation, we adopted the classification of the GTD in this matter. Its definition is that multiple bombs had to detonate both in the same geographical location and at the same time to constitute a single attack. If either timing or geographical location differed between bombs, the incidents are counted as separate attacks.

$N_{i,t} = \sum_{\tau=1}^{t} \sum_{j=1}^{n_\tau} \frac{\mathbf{I}_{i \in A_{j,\tau}}}{|A_{j,\tau}|}$. Using this definition, we describe the averaged reinforcement model as

$$Q_{i,t}^{average} = (1 + N_{i,t-1})^{-1} \left(q + \sum_{\tau=1}^{t} \sum_{j=1}^{n_\tau} \mathbf{I}_{i \in A_{j,\tau}} \frac{y_{j,\tau}}{|A_{j,\tau}|} \right). \tag{2}$$

The two models differ on two accounts in their behavioral implications. Both models incorporate the law of effect (Thorndike, 1898). It states that if a chosen action rendered a successful outcome in the past, the likelihood of choosing the same action in the future is increased. The models differ, however, on their definition of success. In the total reinforcement model, an additional attack on target i will always weakly increase the reinforcement for that target, because any payoff larger than 0 will be counted as a success. Therefore, the probability of attacking a target in the future will, *ceteris paribus*, weakly increase with each additional attack on that target. This is not the case in the average reinforcement model. In this latter approach, any attack that renders a payoff below the historic average of attacks on the same target will decrease the reinforcement and thus the choice probability for that target. The averaging model thus already sets something akin to an aspiration level.

The second dimension in which the two models differ behaviorally is how they assume the reinforcement to be increased over time. In the total reinforcement model, a payoff in an earlier period will increase the absolute reinforcement of the target type by the same amount as an attack in a later period. In the average reinforcement model the weight of each attack in the reinforcement is $(1 + N_{i,t-1})^{-1}$, a number that decreases with each additional attack. In this model, the absolute change in reinforcement due to a successful attack will thus, *ceteris paribus*, decrease over time. Therefore, only the average reinforcement model incorporates the power law of practice, which states that learning curves are concave in accumulated experience (see Newell and Rosenbloom, 1981, for a review).

3.1.2 Choice Probabilities

There are four alternative decision rules that propose how reinforcements can be turned into choice probabilities. The simplest possible rule is to choose the action that has the highest reinforcement with some fixed probability and all other actions with equal probability. This so called *fixed error* model has received only very limited support in the past (Stott, 2006) and will thus be ignored here. The second rule, the *probit* model, relies on the difference in reinforcement and equates the choice probability with the cumulative distribution function of an $N(0, \lambda)$ distribution at the point of the difference. This approach is usually applied to situations in which only two choice alternatives exist (see, e.g., Cheung and Friedman, 1997) and generalizations to more than two actions are not straightforward. For this reason, we will ignore this model here as well.

The two remaining decision rules which we do contrast here are the *logit* model and the *power*

model. The logit model is, for example, used in Camerer and Ho (1999). It is described as

$$p_{i,t} = \frac{e^{\lambda \cdot Q_{i,t}}}{\sum_{l=1}^{I} e^{\lambda \cdot Q_{l,t}}}. \tag{3}$$

The parameter λ can be interpreted as an inverse error term or a measure of sensitivity. Increased sensitivity implies stronger learning from reinforcements, while a decrease in the parameter can either be a sign of less precise learning or simply of mistakes made by the organization. Leaving λ as an open parameter also makes the model scale insensitive. This is a highly desirable property in our analysis, since we do not know the absolute scale of the public attention proxy.

The power model was introduced by Luce (1959) and is, for example, used in Erev and Roth (1998). It is described by

$$p_{i,t} = \frac{(Q_{i,t})^{\lambda}}{\sum_{l=1}^{I} (Q_{l,t})^{\lambda}}. \tag{4}$$

Here, λ has a similar interpretation as in the previous model. As in equation (3), the power model is scale insensitive. Multiplication of all reinforcements by a constant will not change the choice probabilities.

The key difference in the learning process of these two decision rules is the reliance on absolute or relative differences in the reinforcements. The logit model relies on the absolute differences. This means that it does not matter for the choice probabilities whether the reinforcements of two actions are $Q_{1,t} = 1$ and $Q_{2,t} = 2$ or $Q_{1,t} = 101$ and $Q_{2,t} = 102$. This property makes the logit model invariant to an absolute term being added to each reinforcement. If one interprets the size of the reinforcement as some measure of experience (as it is at least the case in the total reinforcement model), the agents' processing of a certain difference in reinforcements does not change when more experience has been accumulated in total. The power model, on the other hand, relies on the relative difference in reinforcement. Thus, the difference in choice probability would be much higher in the case of $Q_{1,t} = 1$ and $Q_{2,t} = 2$ than in the case of $Q_{1,t} = 101$ and $Q_{2,t} = 102$, because the agent corrects a difference in the reinforcements for two actions by the amount of overall experience gained. Behaviorally, this difference between the two model implies that the power model adheres to the power law of practice, while the logit model does not.

There are also operational differences between the two models. The logit model does allow for negative reinforcements, while the power model does not. This is of no consequence in the current analysis, because our measurement of attack payoffs only has positive values. It will, however, matter when analyzing aspiration levels as discussed below. Additionally, only the logit model is indifferent towards adding an absolute term. This makes us unable to identify q when combining total reinforcements with logit choice probabilities. The parameter can only be identified when analyzing the power model.

3.1.3 Empirical Strategy

We analyze the target choices of terrorist organizations for their bombing attacks. Given the choices and payoffs that we can observe, we calibrate four models with three different payoff structures each. The four models are the possible combinations of the total and average reinforcement model with the logit and power choice probability model. The three goal structures are the fatalities, the relative number of Google searches, and a weighted combination of both. To make the payoffs comparable, we normalize both fatalities and public attention proxy for each group $g \in \{1, ..., G\}$ to be in the interval $[0, 100]$ by setting $\bar{y}_{j,t,g}^{fatalities} = \frac{100 \cdot y_{j,t,g}^{fatalities}}{max_{j,t}\{y_{j,t,g}^{fatalities}\}}$ and similarly setting $\bar{y}_{j,t,g}^{p.a.} = \frac{100 \cdot y_{j,t,g}^{p.a.}}{max_{j,t}\{y_{j,t,g}^{p.a.}\}}$. The weighted combination payoff of an attack is then calculated as $y_{j,t,g} = \beta \bar{y}_{j,t,g}^{fatalities} + (1-\beta)\bar{y}_{j,t,g}^{p.a.}$ whereas β is an open parameter calibrated from the data. The normalization of the payoffs is necessary, because the public attention measure already is normalized and cannot be accessed in raw form. The procedure proposed here seems sensible but does lead to a certain temporal inconsistency because payoffs for attacks can be influenced by payoffs of attacks in the future. We address this issue further in Section 5.

For each goal structure and each model, we calibrate all open parameters by maximizing the log-likelihood over the entire sample. Let θ be the set of all open parameters and $t_g \in \{1, ..., T_g\}$ count all periods with attacks of group g. The maximization then reads[4]

$$\max_{\theta} \left\{ LL = \sum_{g=1}^{G} \sum_{t_g=1}^{T_g} \sum_{j=1}^{n_{t_g,g}} ln \left(\sum_{i \in A_{j,t_g,g}} p_{i,t_g,g} \right) \right\}. \tag{5}$$

Note that if an attack features multiple target types, we count the logarithm of the sum of all applicable choice probabilities towards the log-likelihood. This is done to reflect that all these target types would have been the correct prediction by the model. The index g in $n_{t_g,g}$ indicates that for each group only its own attacks are counted towards its log likelihood.

We then compare the fit of the four models within each payoff structure. As shown in Table 2, the sample sizes differ between the different payoff measures due to data availability. A comparison between different payoff structures is thus generally not meaningful. The models we calibrate feature different numbers of free parameters. If two models are nested within each other, the inclusion of an additional free parameter will always weakly increase the log-likelihood, but this is not necessarily an indication of a better fit due to the possibility of overfitting. We thus use the standard ways of adjusting the in-sample fit for additional parameters and report both the Bayesian information criterion ($BIC = (k) \cdot ln(M) - 2LL$, with k being the number of free parameters and M being the total number of observations in the sample) and the Akaike information criterion ($AIC = 2k - 2LL$). To provide a frame of reference, we also report the log-likelihood of a maximum entropy model, which assumes equal choice probabilities for all target types in all periods.

It can be argued that the in-sample fit is not particularly informative of the actual fit and

[4]Further details on estimation method and data preparation can be found in the Online Appendix.

that both the BIC and the AIC do not sufficiently adjust for the number of free parameters. An alternative approach is to report an out-of-sample fit. To ensure that we do not overfit the data, we thus also report the out-of-sample fit of our models and focus our discussion on these results, employing the same procedure as Camerer and Ho (1999). For each terrorist organization, we calibrate the decision model for all $t_g < 0.7T_g$ then report the fit of the calibrated model in the remaining time periods.[5] This procedure assumes that parameters are constant over time. While this assumption can certainly be argued, it seems more realistic in our application, in which the environment changes over time and adaptation is essential for the terrorist organization, than in applications in the laboratory, where the steady state is often reached after some time and adaptation ceases to be important.

We measure the out-of-sample fit with two different proper scoring rules. We first report the log-likelihood and compare it to the log-likelihood of the maximum entropy model. This is done by constructing the measure $\rho^2 = \frac{LL^{max.entropy} - LL^{model}}{LL^{max.entropy}}$ which shows how much more the model is able to explain compared to the uninformative maximum entropy alternative. Since the logarithmic nature of the log-likelihood punishes small probabilities of selected actions quite severely (Selten, 1998), we also report the mean squared deviation (MSD) of the choices from the choice probabilities in the non-calibration part of the sample. If M^{oos} is the total number of out-of-sample observations over all groups, the MSD is defined as

$$MSD = (M^{oos})^{-1} \sum_{g=1}^{G} \sum_{t \geq 0.7 \cdot T_g}^{T_g} \sum_{j=1}^{n_{t,g}} \sum_{i=1}^{I} (p_{i,t,g} - \mathbf{I}_{i \in A_{j,t,g}})^2. \tag{6}$$

3.1.4 Results of the Model Selection Analysis

The results of the initial model selection analysis are displayed in Table 3. Overall, the model fit is very reasonable. The best model, according to out-of-sample fit, lies above 18% in the ρ^2 measure. Comparing these results to the laboratory data on games studied in Camerer and Ho (1999), we find that we can explain the target choices of terrorist organizations about equally well. Such comparisons, however, always need to be interpreted with great care, because the analyses are made on very different data.

The results show that for every goal structure, the combination of total reinforcement model with power choice probabilities has the highest explanatory value. This model's improvement over equal probabilities is highly statistically significant for all goal structures when using a likelihood ratio test. This is true for all measures of fit, be it in-sample or out-of-sample.[6] With respect to the learning model, this can be interpreted as evidence against the power law of practice, against an aspiration level, or a combination of both. At the same time, however, the total reinforcement

[5] We conducted a robustness analysis on the size of the calibration set, ranging it from 0.3Tg to 0.8Tg (see Online Appendix G). Our findings do not change materially with this variation.

[6] For the remainder of the paper, all reported improvements on nested models are significant at the 1% level unless reported otherwise.

Table 3 – Results of Model Selection Analysis

Models			In-Sample			Validation		
$Q_{i,t}$	$p_{i,t}$	k	LL	AIC	BIC	LL	ρ^2	MSD
Payoff: Fatalities (M = 20,170; G = 69)								
Max Entropy		0	-31664	63327	63327	-9526	-	16.96%
Total	Logit	1	-30323	60647	60655	-9006	5.45%	16.19%
Average	Logit	2	-31664	63331	63347	-9527	-0.02%	16.96%
Total	Power	2	*-29756*	*59516*	*59531*	*-8785*	*7.78%*	*15.85%*
Average	Power	2	-31664	63331	63347	-9526	0.00%	16.96%
Payoff: Public Attention (M = 8,147; G = 28)								
Max Entropy		0	-12454	24907	24907	-4127	-	17.46%
Total	Logit	1	-11199	22400	22407	-3730	9.63%	16.29%
Average	Logit	2	-12454	24911	24925	-4155	-0.68%	17.57%
Total	Power	2	*-10317*	*20637*	*20651*	*-3378*	*18.16%*	*14.89%*
Average	Power	2	-12454	24911	24925	-4127	0.00%	17.46%
Payoff: Fatalities and Public Attention (M = 7,928; G = 28)								
Max Entropy		0	-12112	24224	24224	-3992	-	17.49%
Total	Logit	2	-10727	21457	21471	-3530	11.57%	15.58%
Average	Logit	3	-12112	24230	24251	-3991	0.02%	17.49%
Total	Power	3	*-10031*	*20068*	*20089*	*-3260*	*18.33%*	*14.88%*
Average	Power	3	-12112	24230	24251	-3992	0.00%	17.49%

The table displays the results of the model selection analysis. Each section of the table reports the analyses for all four possible combinations of reinforcement model (indicated in column $Q_{i,t}$) and choice probability model (indicated in column $p_{i,t}$). The third column reports the number of estimated parameters. The first three columns of results report the in-sample fit of each model. The second three columns report the out-of-sample fit when the calibration sample is comprised of the first 70% of observations for each terrorist organization and the validation sample is the last 30% of observations. The samples for the three horizontal sections differ due to the assumed payoff. The best fitting model for each sample is highlighted in italics.

model performs much better with power choice probabilities, which incorporate the power law of practice, than with logit choice probabilities, which do not incorporate this motive. As such, the best fitting model does feature diminishing marginal sensitivity in learning with growing experience, but no aspiration level concept. This result seems congruent with other findings in the literature. Evidence for the power law of learning is a robust finding in an abundance of studies (see, e.g., Newell and Rosenbloom, 1981, for a review). The lacking support for an aspiration level can be seen as part of a broader discussion of mechanisms in organizational learning. Early literature, beginning with Cyert and March (1963), highlights the importance of an aspiration level in organizational learning and initial support for this hypothesis has been gathered from legitimate organizations[7] (e.g., Greve, 1998; Miller and Chen, 2004). However, these results have more recently been called into question by Denrell (2008) who shows that what has been taken as evidence for aspiration-based behavior might simply be variable risk preferences of organizations. Our analysis supports

[7]That is organizations acting within legal bounds and not engaged in violent conflict as the organizations studied here.

this contention by showing that a direct estimation of the learning process favors models without an aspiration level concept.

We can see that the average reinforcement model performs poorly. It never performs better than the maximum entropy model in-sample, regardless of payoff structure. Furthermore, the model is calibrated such that it resembles the maximum entropy model in-sample (this can be reached by either calibrating q to be very large or λ to be close to zero). For logit choice probabilities, it has a slightly different out-of-sample fit than the maximum entropy model, but it performs worse than the benchmark in the first two analyses and only very slightly, and statistically insignificantly (D = 1.3856, p-value = 0.7089), improves the out-of-sample fit over the uninformative case in the model with both payoffs. We thus conclude that the average reinforcement model holds no information about the behavior of terrorist organizations.

Next to the good performance of total reinforcement and power probabilities, we can also observe the out-of-sample fit to be best for the analyses including both payoffs. In particular, it is much better than the analysis only considering fatalities. This result most likely has two reasons. Firstly, fatalities might not be the main objective for all terrorist organizations. Asal and Rethemeyer (2008) find attacks from religious and ethnonationalistic organizations to be more lethal than those of leftist organizations, pointing towards an ideology driven difference in the goal structure. An analysis which only considers fatalities as payoffs simply might not capture the correct measure of reinforcement for some organizations. Secondly, the data quality of the GTD is relatively low for the earlier decades. Data for the year 1993 is only sparsely available and many incidents in the 1970s and 1980s are not fully covered. The public attention measure restricts the analysis to start in the year 2004 where data quality is much higher. Fatalities nevertheless add to the explanatory power. The out-of-sample fit is better if the goal structure is taken as dichotomous rather than solely focused on public attention. This result becomes more pronounced when only considering the 7,928 observations for which data on both payoffs are available. Such a result also allows for a likelihood ratio test which reports a highly significant improvement (D = 80.7774, p-value < 0.01). Detailed results for such an analysis can be found in the Online Appendix. Given these results, we concentrate all further analyses on this goal structure.

3.2 Behavioral Motives

In this section, we analyze two commonly suggested behavioral motives for reinforcement learning: information discounting and endogenous aspirations. Our discussion is based on the total reinforcement model with power choice probabilities which showed the best fit in the previous section's analysis. We also repeated the analysis on behavioral motives using logit choice probabilities. Neither the in-sample nor the out-of-sample fit of these estimations were as good as the basic model with power choice probabilities (see the Online Appendix for details).

3.2.1 Model Set-up

The total reinforcement model that includes information discounting and endogenous aspirations can be described as $Q_{i,1} = q$ in the first period and for all following periods as

$$\begin{aligned} Q_{i,t} &= \delta^{\Gamma(t)} Q_{i,t-1} + \sum_{j=1}^{n_{t-1}} \Big[\mathbf{I}_{i \in A_{j,t-1}} \mathbf{I}_{y_{j,t-1} \geq Asp_{t-1}} \frac{y_{j,t-1} - Asp_{t-1}}{|A_{j,t-1}|} \\ &\quad + \mathbf{I}_{i \notin A_{j,t-1}} \mathbf{I}_{y_{j,t-1} < Asp_{t-1}} \frac{Asp_{t-1} - y_{j,t-1}}{I - |A_{j,t-1}|} \Big] \text{ for } t \geq 2. \end{aligned} \tag{7}$$

Information discounting appears when the discounting factor δ is unequal one. The discounting factor is taken to the power of $\Gamma(t)$ which governs whether discounting is based on the number of attacks or the number of days as is explained below. Endogenous aspirations appear in (7) as the term Asp. They are time dependent and endogenously formed from prior reinforcements. The progression of the aspiration level starts with $Asp_0 = b$ and proceeds as

$$Asp_t = \alpha Asp_{t-1} + (1 - \alpha) \sum_{j=1}^{n_{t-1}} \frac{y_{j,t-1}}{n_{t-1}} \text{ for } t \geq 1. \tag{8}$$

While b thus denotes initial aspirations, $\alpha \in [0, 1]$ determines the speed with which aspirations adjust to recent reinforcements. Given that $y_{j,t,g} \in [0, 100]$, the aspiration level has no impact on target choices if $b = 0$ and $\alpha = 1$.

We will focus on an analysis using both public attention and fatalities as payoffs. As such, the payoff for each attack is given by $y_{j,t,g} = \beta \bar{y}_{j,t,g}^{fatalities} + (1 - \beta) \bar{y}_{j,t,g}^{p.a.}$ as described above. The probabilities are formed according to the power model given in equation (4). By focusing on power probabilities due to their descriptive superiority to the logit model, we are using a model that does not allow for negative reinforcements. However, if an aspiration level is included in the analysis, performance below aspirations needs to lead to a decrease in choice probability. To achieve this without potentially requiring negative reinforcements, we divide the negative performance by the number of non-chosen alternatives and add the amount to the reinforcements of all other target types (the second half of the sum in equation (7)). This decreases the choice probabilities of the chosen target types if they performed below aspirations and, at the same time, leads to the sum of reinforcements being increasing in t, which is in accordance to the power law of practice.

The full model has 6 free parameters: α, β, δ, λ, b and q. Note that in equation (7) both information discounting and aspiration updating are nested. That is, they disappear if $\delta = 1$ or $b = 0 \wedge \alpha = 1$, respectively. We can thus estimate both behavioral motives separately or in combination when using equation (7). A comparison of the fit will then inform us about which factors drive the behavior of terrorist organizations. Before this is done, however, we will interpret the behavioral implications of the two motives.

3.2.2 Behavioral Implications

The common assumption of *belief discounting* is featured in several models of learning in game theoretic environments (Erev and Roth, 1998; Camerer and Ho, 1999). The idea is that information gained from more recent actions impacts estimates about future actions more than information gained from actions further in the past. Information discounting in target choices of terrorist organizations is very likely. The circumstances in which terrorist organizations operate change constantly. The current societal climate will influence the public attention given to a terrorist organization due to an attack on a certain type of target. Further, changes in infrastructure or technology can influence how many people will be killed by a bomb attack on a certain target. For example, some cities changed the way they build bus stops after the 9/11 attacks, making shrapnel fatalities after bombings of bus stops much less likely. Lastly, and probably most importantly, the weapons technology of terrorist organizations has markedly changed over our observation period. The broad availability of cheap cell phones, for example, makes detonating bombs remotely much easier now than it used to be in the past.[8] Belief discounting captures terrorist organizations' reactions to such changes in our model by slowly decreasing the importance of attacks as time proceeds. Mathematically, we implement this motive by assuming the reinforcement of the last period to be discounted by the factor $\delta^{\Gamma(t)}$.

There are two possible ways discounting can appear in our empirical setting. One can assume that the terrorist organization sees each committed attack as one period and thus discounts based on how many attacks ago a payoff was observed. In this case, we set $\Gamma(t) = 1$ for all t. Alternatively, the terrorist organization can be assumed as sensitive to the time which has passed between two attacks. If that is the case, the organization discounts more strongly if more time has passed between two attacks than if they are following each other closely. This second case is modeled by setting $\Gamma(t)$ to be equal to the number of days between the attack in period $t-1$ and the attack in period t. The question is ultimately an empirical one. We thus estimate both models and see which one has the better in-sample and out-of-sample fit.

Over repeated iterations of the model, the discounting factor is applied to a payoff multiple times. Consider, for ease of exposition, a model with $b = 0$ and $\alpha = 1$. The reinforcement for target i in $t = 4$ if the first three attacks were on target i will then be: $Q_{i,4} = \delta^{\sum_{\tau=1}^{3}\Gamma(t)}q + \delta^{\sum_{\tau=2}^{3}\Gamma(t)}y_{t=1} + \delta^{\Gamma(3)}y_{t=2} + y_{t=3}$. The weight of a payoff thus decreases hyperbolically over time if $\delta < 1$. We do not implement this parameter restriction ex-ante. As such, if discounting does not play a role, the data could lead to $\delta = 1$ or even $\delta > 1$ if the terrorist organization is placing less weight on payoffs from more recent attacks than on those from attacks further in the past.

The notion of *aspiration levels* was first introduced to economics by Friedman and Savage (1948) and further formalized and made popular by the works of Cyert and March (1963) and Kahneman and Tversky (1979). In reinforcement learning applications, the idea is closely linked to the concept of problemistic search. If an action leads to an unsatisfactory outcome (i.e. an outcome

[8] We thank an anonymous reviewer for pointing out this particularly striking example of technological change

below the aspiration level), the agent is going to search for alternative options by increasing the choice probabilities of alternative actions. This is in stark contrast to the basic form of the total reinforcement model in which every prior action can only increase the reinforcement and thus the choice likelihood of that action in the future.

The question of how aspiration levels are formed is still an open issue (e.g., Köszegi and Rabin, 2006; Baucells et al., 2011; Sprenger, 2016). It does, however, seem reasonable to assume in our application that the aspiration level is not exogenously given, but rather determined by the experience of the decision maker. We thus opt for the recursive updating mechanism given in equation (8). This has received empirical validity in experimental studies (Lant, 1992, see, however, Baucells et al. 2011 for contrary evidence) and has formerly been applied in learning models for individual decision making (Börgers and Sarin, 2000), organizational learning (Jaspersen and Peter, 2017) and in different OR models (e.g., Popescu and Wu, 2007; Wu et al., 2015).

3.2.3 Results of the Behavioral Motives Analysis

The results of our estimations are shown in Table 4. We find empirical evidence for belief discounting in the behavior of the terrorist organization. Both in-sample and out-of-sample fit is increased if this motive is added to the basic model. This increase is highly statistically significant when performing a likelihood ratio test. Among the discounting models, we see more support for attack based discounting ($\Gamma(t) = 1$ for all t) than for time based discounting. Even though little research on the question whether to discount by trial or to discount by time passed has been conducted so far, this result nevertheless seems supported by prior literature. If we consider that belief discounting has primarily been documented in laboratory studies (e.g., Camerer and Ho, 1999; Hertwig et al., 2004), environments in which very little time passes between trials, then discounting by trial seems to be the more likely model than discounting by time to explain these findings.

We cannot find any support for aspiration levels to play a role in the decision process. Adding this motive to the basic model increases neither in-sample nor out-of-sample fit in an economically or statistically meaningful way. Furthermore, the parameters calibrate towards $b = 0$ and $\alpha = 1$. As such, we find no support for target payoff levels in our data. The best fitting model includes only the power law of practice (through the power choice probabilities) and discounting. Lacking evidence for the aspiration levels is in line with the results reported in Table 3, that is the rejection of the average reinforcement model in favor of the total reinforcement. As we argued above, this result is congruent with recent developments in organizational learning theory and behavioral adaptation.

It could be argued, however, that we did not implement aspiration levels as it is commonly done in theoretical analyses (Cross, 1973; Börgers and Sarin, 2000; Jaspersen and Peter, 2017). In such studies, probability updating from reinforcement is indirect in the sense that probabilities are transformed only on the basis of comparing the last payoff with the aspiration level. To make sure that it is not a simple model misspecification which drives our results, we also estimate such an indirect aspirations model. Details of the model and the empirical assumptions are given in the Online Appendix. As shown in the last horizontal section in Table 4 the indirect model has poorer

Table 4 – Results of Behavioral Motives Analysis

Model	k	In-Sample			Validation		
		LL	AIC	BIC	VLL	ρ^2	MSD
Benchmark							
Max Entropy	0	-12112	24224	24224	-3992	-	17.49%
Behavioral Motives							
Basic Model	3	-10031	20068	20089	-3260	18.33%	14.88%
Discounting (attack)	4	*-9788*	*19584*	*19612*	*-3187*	*20.16%*	*14.47%*
Discounting (days)	4	-9834	19676	19704	-3208	19.64%	14.51%
Aspiration Level	5	-10031	20072	20107	-3260	18.33%	14.88%
Both (attacks)	6	-9788	19588	19630	-3187	20.16%	14.47%
Both (days)	6	-9834	19680	19722	-3208	19.64%	14.51%
Additional Model							
Indirect Asp.	4	-11332	22673	22701	-3523	11.75%	15.77%

The table displays the results of the behavioral motives analysis. The first horizontal section reports the benchmark of the maximum entropy model. In the second horizontal section, the first column of the table lists the behavioral motives which are included in the model. The second column reports the number of estimated parameters. The third horizontal section reports an additionally analyzed model. The first three columns of results report the in-sample fit of each model. The second three columns report the out-of-sample fit when the calibration sample is comprised of the first 70% of observations for each terrorist organization and the validation sample is the last 30% of observations. The best fitting model is highlighted in italics for each statistic.

fit with the data than the reinforcement learning model. This can be considered as a positive result for the direct reinforcement model as well as further evidence against aspiration levels in modeling the behavior of terrorist organizations.

The parameters of the best fitting model of our behavioral analysis are given in Table 5 below. Parameters reported are those from the calibration sample. Standard errors are provided through Jackknifing. For this, we estimate the parameters G times, each time leaving out one terrorist organization. The standard errors of each parameter (here we use β as an example) are then calculated as $\sigma(\hat{\beta}) = \left[(G-1)G^{-1}\sum_{g=1}^{G}(\hat{\beta}_g - \bar{\hat{\beta}})^2\right]^{\frac{1}{2}}$.

The learning model is calibrated with high precision (or, equivalently, with little error) because λ is calibrated close to 1 and has little variation. The learning speed, inversely measured by q, is quite small in comparison to the reinforcements which could range from 0 to 100. However, this estimate has considerable variation attached to it, pointing towards different learning speeds for different organizations. On average, the goal structure gives about three to seven times as much weight to public attention than to fatalities, but this estimate has a very large standard error. In line with Asal and Rethemeyer (2008) it can thus safely be assumed that the diverse organizations have widely differing goal structures. Belief discounting behaves as we had speculated. The discounting factor is below 1 and has a small standard error, such that behavior conforming with discounting seems to be a general characteristic of terrorist organizations.

Table 5 – Calibrated Parameters in Behavioral Motives Analysis

	q	λ	β	δ	α	b
Basic Model	33.4817	1.0995	0.2064	-	-	-
	29.2441	*0.1955*	*0.5194*	-	-	-
Discounting (attacks)	25.5434	0.9128	0.1230	0.9647	-	-
	15.2821	*0.0378*	*0.1320*	*0.0079*	-	-
Discounting (days)	38.9182	0.9601	0.1686	0.9975	-	-
	49.7994	*0.0457*	*0.0887*	*0.0007*	-	-
Aspiration Level	33.7313	1.1094	0.2565	-	0.0000	1.0000
	29.0747	*0.1795*	*0.2108*	-	*0.0000*	*0.0000*
Both (attacks)	25.5432	0.9128	0.1230	0.9647	0.0000	1.0000
	15.2821	*0.0378*	*0.1320*	*0.0079*	*0.0000*	*0.0000*
Both (days)	38.9180	0.9601	0.1686	0.9975	0.0000	1.0000
	49.8019	*0.0457*	*0.0887*	*0.0007*	*0.0000*	*0.0000*

The table displays the results of the calibrated model described in equation (7) with different restrictions which equal those of the models displayed in Table 4. The reported coefficients are estimated from the calibration sample comprised of the first 70% of observations for each group. Standard errors are Jackknifed and displayed in italics.

4 Predictions across Groups

A compelling feature of reinforcement learning models is that they can be applied to newly emerging agents without any history of activity. We test in this section whether the model analyzed here can be used in this way, so policy makers could employ our model to cope with nascent terrorist groups. For this analysis, we change the out-of-sample validation procedure. We first calibrate the model on data of all but one organization and then analyze the fit of this model for the target choices of the excluded organization, to assess to which extent the model could predict its adaptive behavior.

The results of this analysis are reported in Table 6. As before, the model including information discounting on the basis of attacks and no aspiration level performs best in the analysis. More interestingly, this model is, on average, able to significantly decrease the prediction error of an equal likelihood model. While the ρ^2 score does not quite reach the level of the within-group predictions, it still reports an increase in log-likelihood by 18.75%.

The findings from this analysis offer some interesting possible applications. They imply that data from organizations following a specific ideology in one part of the world can be applied to predict the actions of another organization, potentially following a different ideology and operating in a different part of the world. To show this, we calibrated the model with data from all organizations except the Liberation Tigers of Tamil Eelam (LTTE) and then applied it to predict the actions of the LTTE. While attacks in our data set are mostly carried out by religiously motivated groups in the Middle East, actions of the LTTE are not primarily motivated by religion. Additionally, the LTTE is the only organization in our data set operating in the country of Sri Lanka.[9]

[9]Note that we did not select the LTTE due to a particularly good fit of the data. In comparison to the equal

Table 6 – Results of Behavioral Motives Analysis across Groups

Model	k	In-Sample			Validation		
		LL	AIC	BIC	VLL	ρ^2	MSD
Benchmark							
Max Entropy	0	-12112	24224	24224	-12112	-	17.49%
Behavioral Motives							
Basic Model	3	-10031	20068	20089	-10126	16.40%	15.08%
Discounting (attack)	4	*-9788*	*19584*	*19612*	*-9842*	*18.75%*	*14.63%*
Discounting (days)	4	-9834	19676	19704	-9918	18.11%	14.73%
Aspiration Level	5	-10031	20072	20107	-10125	16.41%	15.06%
Both (attacks)	6	-9788	19588	19630	-9842	18.75%	14.63%
Both (days)	6	-9834	19680	19722	-9918	18.11%	14.73%

The table displays the results of the behavioral motives analysis when the validation sample is a randomly picked group. The first horizontal section reports the benchmark of the maximum entropy model. In the second horizontal section, the first column of the table lists the behavioral motives which are included in the model. The second column reports the number of estimated parameters. The first three columns of results report the in-sample fit of each model and are by construction equal to the corresponding columns in Table 4. The second three columns report the out-of-sample fit. Here, the analysis was repeated 28 times. In each repetition, 27 groups acted as the calibration sample and one group was used as the validation sample. Fit measures are reported in the aggregate over all 28 repetitions. The best fitting model is highlighted in italics for each statistic.

Results of this specific application are shown in Figure 1 where we summarize the empirical and predicted likelihoods of attack on an annual basis. The model shows a good fit in a visual inspection. Counterterrorism agencies could use the model reported as best fitting here to predict attack patterns of emerging organizations and deploy countermeasures accordingly. Even with the coarse publicly available data employed here, the fit of the model already looks promising. Additional data, which might be available to counterterrorism agencies, would likely further increase its predictive accuracy.

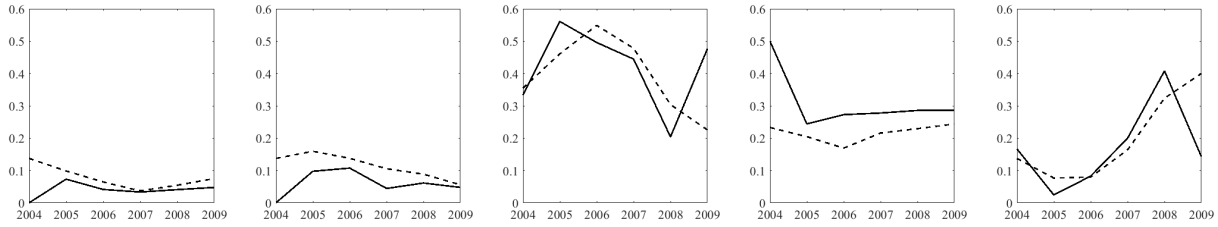
5 Limitations, Robustness and Alternative Models

The described estimation procedures imply several assumptions which need to be discussed. It could be a concern that our result on aspiration levels stems from the fact that they are the only behavioral motive which can possibly imply negative reinforcements of an attack. Since there is no standard assumption on how negative reinforcements influence choice probabilities in the power model, our way of modeling it could be the reason for the bad performance of aspiration levels. However, when repeating the analysis with logit choice probabilities, which can accommodate negative reinforcements, results of our behavioral motives analysis do not change. As for all other robustness checks reported below, results for this analysis can be found in the Online Appendix.

It is obvious from equation (5) that we treat the decision making process of all terrorist organiza-

probability model, using reinforcement learning decreases the log-likelihood of their attacks by about 19%, which is approximately equal to thee average reduction of the model reported in Table 6.

Figure 1 – The figure shows the empirically observed (solid line) and predicted (dashed line) likelihoods of bombing attacks on different target types by the Liberation Tigers of Tamil Eelam (LTTE) in their active time of operation after 2004. The model used for prediction was the best fitting model from the analysis reported in Table 6 calibrated on the data of all other terrorist organizations, except those of the LTTE. The year 2010 was ignored because it only featured two attacks by LTTE and thus did not allow for reliable calculation of empirical probabilities.



(a) Business (b) Government (c) Military (d) Private Citizens (e) Infrastructure

tions as homogeneous. The reason is our aim of drawing general insights on terrorist organizations' adaptive behavior. Additionally, individual level estimations of reinforcement learning models often lead to widely varying parameters (Cheung and Friedman, 1997) and thus the estimation of homogeneous models is common in the literature (Camerer and Ho, 1999). Nevertheless, when calibrating the models for each group separately, reinforcement learning commonly improves on a maximum entropy model and the behavioral motive of discounting is more prevalent than that of an aspiration level.

We also implicitly assume all attacks to be equally costly. This is certainly a simplification, since there are harder and softer targets (Berman and Laitin, 2008). The degree of simplification is somewhat lessened by the fact that our analysis focuses on bomb attacks alone. While there will be differences in costs between different bomb attacks, these differences will be less pronounced than in an analysis across all different attack types. Allowing for heterogeneous costs between different target types (which stay constant over time) does not change our results on behavioral motives. Our results thus seem to be independent of the assumption of equal costs for all target types.

When considering assumptions about cost structure, we further need to contrast our approach with those based on cost–benefit analyses. Recently, Hausken (2018) provided an in-depth analysis of possible motivating benefits as well as the associated costs of an attack from a terrorist perspective. Regarding benefits, he makes the case for measuring human damages (such as fatalities), gained influence (by demonstrating the ability to carry out an attack) and economic damages (such as destroyed property). Our pay-off proxies of fatalities and public attention do not fully capture the first two of his proposed benefits, but are similar in their motivation. Economic damages are hard to measure in our data and are thus not part of our analysis, even though they would likely increase the predictive accuracy of our models if data on them were obtainable. The crucial difference between our approach here and Hausken's (2018) proposition is his inclusion of detailed and attack-specific costs in three similar categories as those used for measuring the benefits. However, such an approach is unfeasible in our setting given the sheer magnitude of the analyzed data. We

see significant potential for future research in the application of detailed cost structures. This is particularly true in case-based analyses of specific organizations, where data collection is more feasible than in our setting.

As is common in reinforcement learning models, we assume that organizations only learn from their own experience. Vicarious learning is excluded from the models. Many of the counter-terrorism models referenced throughout the paper focus on interaction of a defender and a single attacker. The assumption is thus in line with our motivation, which is focused on developing models of behavioral adaptation for such applications. However, there is at least anecdotal evidence that some learning between organizations takes place (Hoffman, 2006). Hence, we included vicarious learning in a robustness check. The results show a strong increase in the fit of the model (ρ^2 increases by 1.25 percentage points) and show the results of our behavioral analysis to be robust. Further investigating the nature of vicarious learning in more detail, such that models for networks of terrorist organizations can be developed, is a promising direction of further research. Preliminary analysis performed with our model specification showed that the intensity of vicarious learning increased with both ideological and geographic proximity. However, a more substantial analysis based on a discussion of the appropriate functional form is necessary to make credible statements about the exact nature of the vicarious learning process.

As noted above, the normalization chosen in the empirical analysis leads to a certain amount of temporal inconsistency. Payoffs at time $t = 1$ can be influenced by payoffs in periods $t > 1$ through the normalization procedure. Unfortunately, the already normalized nature of the public attention payoff proxy makes it impossible for us to use no normalization in our analysis. To analyze whether the temporal inconsistency biases our interpretations, we repeated the analysis with a different normalization which is more temporally consistent at the cost of restricting the behavioral realism of the model. A detailed description of the procedure and the results is given in Section E.6 of the Online Appendix. Results are consistent with those reported in the main analysis making us confident that our interpretations are not biased due to the temporal consistency issue.

We restrict the sample to bombings to ensure the highest degree of comparability between attacks. This implies the assumption that terrorist organizations first make a choice about their mode of attack and then choose the target. However, it is entirely possible and in some cases sensible that terrorist organizations choose a target first and select the most appropriate mode of attack after the target selection. To get an indication about how much our results depend on the implicit assumption about the decision process, we repeated the analysis without the sole focus on bomb attacks and instead considered both bombings and attacks categorized as armed assaults in the GTD. Together, these two attack modes comprise over 82% of attacks since 2004. The logic behind this analysis is that we now treat the choice of target type as independent of the choice of attack mode, assuming that the terrorist organizations learn about target types across different attack modes. The inclusion of the additional attacks strongly increased the out-of-sample fit of our model (ρ^2 increases by almost 7 percentage points) but otherwise left the result of the behavioral model analysis unchanged. This result indicates that our conclusions can be drawn

independently from the initial assumption about the order of the decision process. Even though the results with more attack modes seem to be even stronger evidence for reinforcement learning by terrorist groups than those reported in our main analysis, we keep our focus on bombings alone. This way, the analyzed actions are more homogeneous between each attack and period and we analyze a decision situation with the least amount of heterogeneity.

Lastly, we make the assumption of equal initial reinforcements of all target types. Alternatively, we could have estimated different initial reinforcements for the targets which are the same for all terrorist organizations. We refrained from doing so, because we deemed the underlying assumption of such an estimation as unrealistic. In the fatalities analysis, we calibrate the reinforcement learning process of terrorist organizations from significantly different periods of recent history. It is unrealistic to assume that the Shining Path movement in the late 1970s and Al-Shabaab in 2007 would have the same relative ordering of attacks on private citizens and attacks on infrastructure. By assuming equal initial reinforcement for all attack types, these organizations can quickly establish their own relative ordering rather than needing to reverse a superimposed one first.

The approach taken here can further be contrasted with other decision models which have been proposed in the literature. Above, we discuss how a cost–benefit analysis of terrorist organizations can be applied to the stochastic learning model we use in our analysis. It is, however, also possible to use such models in non-stochastic models or models not based on reinforcement. Hausken (2018) introduces his model in a forward-looking, deterministic setting where possible consequences and their probabilities are assessed in advance and an attack is carried out if its expected utility is positive. For yet another alternative use of cost–benefit analysis in a stochastic model, see Hausken and Moxnes (2001) in which a stochastic model is used to predict frequencies of actions. The model is similar to ours in the sense that larger expected (net) payoff of an action leads to more actions, but places no restrictions on the process generating costs and benefits, while we assume that the reinforcement of an action is created from historic payoffs only.

6 Discussion, Implications and Conclusions

The identification of optimal strategies for countering terrorism is an important topic of military and intelligence OR. Prior studies rely on analyzing dynamic systems (e.g., Kress and Szechtman, 2009) or game theoretic equilibria (e.g., Zhuang and Bier, 2007). In this paper, we propose an additional tool for such analyses which is based on understanding the learning patterns and the behavioral adaptation of terrorist organizations while they aim to achieve multiple objectives. This tool can both be used as an assumption about adaptive behavior in dynamic models as well as in analyses of the convergence phase in a repeated game. Based on data of terrorist organizations' target choices, we show that a model of total reinforcement learning with power choice probabilities and belief discounting has the best fit to describe the observed adaptive behavior. We calibrate the parameters of the model based on the data and thus provide a possible starting parameterization for future research. In addition to the implications for future modeling efforts, our results also offer some immediate insights into the adaptive behavior of terrorist organizations and corresponding

prescriptive advice for counter-terrorism strategies. These are discussed below.

This study is concerned with tactical counter actions, not overarching strategic concerns in handling terrorism on a societal level. Other studies, such as Glaeser (2005) and Bier and Hausken (2011) focus on more macro-level questions of preventing hatred and conflict to appear in the first place or with deterring terrorists from attacking through positive or negative incentives. This paper takes the terrorist organizations' attacks as given and aims to model the adaptive behavior that may explain target choices in the attack.

One of the primary purposes of this type of tactical analysis is to utilize the model for predicting where organizations will strike next. One can think of different applications where this could be useful. Several Arabian terrorist organizations, for example, utilize kidnappings both as a way to increase their renown and as a source of income. Identifying different possible kidnapping victims (such as local politicians, military figures, foreigners, etc.) and evaluating the reinforcements previously gained from taking them hostage could provide data that can be utilized in a model of reinforcement learning and adaptive behavior to make (stochastic) predictions about future targets of kidnappings by these groups.

The behavioral motives represented by the different elements of the best fitting model in our analysis also provide some insights for counter-terrorism strategy. Our results show that terrorist organizations seem to learn about their targets according to the power law of learning. This implies that they have steeper learning curves in the beginning of their existence than they do later on. Terrorist organizations will thus exhibit more inertia in their modus operandi the longer they are in operation. This is an aspect of their behavior which could be exploited by counter-terrorism strategists. It implies terrorist organizations to be more creative in the beginning of their organizational lives than towards the end.[10] Similarly, the organizations are less likely to adapt to new environments if they have been in existence for a longer time. Countermeasures against such organizations should thus be timed to coincide with changes in environment. This does need not be limited to defense strategies of specific targets. The knowledge that terrorist groups display inertia could also allow governments to design countermeasures on a more fundamental level. The two Syrian insurgent groups al Nusra Front and ISIL, for example, differed on the extremism with which they enforced Wahhabist ideology among their followers (Neumann, 2016). Which of their respective strategies was more successful depended, to a large extent, on the circumstances they operated in. Inertia could, should these circumstances change, prevent them from adjusting their strategy. This would be an ideal moment for propaganda campaigns to undermine the insurgents' connection to the population of the territory they control. Carrying this logic one step further, governments could also try to change the environment themselves if they judge the inertia of a hostile organization to be high enough.

[10]This consideration requires a discussion about what constitutes the birth of a terrorist organization. Groups can originate from new and individual circumstances (e.g., the Red Army Fraction in West Germany of the late 1960s) or split off from another group (e.g., the Salafist Group of Preaching and Combat in Algeria in the late 1990s). In our analysis, we treat an organization as new regardless of the circumstances of its origin. However, further research is required to test the validity of this assumption.

The data also show more recent observations to be more important for explaining the behavior of terrorist organizations than those further in the past. This should be considered when extrapolating future attack patterns from past attacks. Even though we do provide evidence for the power law of practice, our results do not imply that the future attack patterns of a given organization must be equal to those from several years ago. Rather, our data implies that the more recent attacks of an organization are good predictors for its next attack. As with the power law of practice, the existence of belief discounting in the adaptive behavior of terrorist organizations highlights the parallels to individual and organizational decision-making in other circumstances. That we can find similar motives in the decision-making of terrorist organizations makes it likely that other results from organizational research and decision analysis may also be informative about terrorist groups. Though we are not the first to use organizational theory to understand terrorist groups (see, e.g., Enders and Jindapon, 2010; Feinstein and Kaplan, 2010), our results corroborate such lines of analysis.

The absence of an aspiration level in the best fitting model tells us that terrorist organizations are not striving for particular target levels in the outcomes of their attacks, but will evaluate both small gains and large gains positively. This is in line with the western operation of ISIL and other radical Islamist organizations who seem to encourage attacks even if the expected gain is likely small (as in the various knife attacks recently observed in Europe). It is natural to distinguish a terrorist groups' attacks into failures and successes. Our results, however, highlight that if an attack was completed, the terrorist organization will count it as a success, almost regardless of the outcome. Further evidence of the absence of an aspiration level would help to solidify our interpretation of the reported empirical results. One avenue to further explore this issue would be to qualitatively analyze communications of terrorists' leadership for the presence or absence of aspirations levels when talking about future attacks.

We base our payoff function on the considerations of Richardson (2006) and model both revenge and renown. That the decision weight β of our payoff function does not calibrate to a value of 0 or 1 implies that both motives do play a role, giving empirical validity to her qualitative argumentation. The significant standard error associated with the estimate of β also implies the value functions to differ between organizations, a finding that is in line with prior results (Asal and Rethemeyer, 2008). Studies of terrorist organizations further make clear that other payoffs than revenge and renown are likely to influence their actions (Richardson, 2006; Siebert et al., 2016; Hausken, 2018). In this study, we limited our analysis to those payoffs which we can measure reliably and consistently for all organizations. Further research is encouraged to consider alternative and additional payoff metrics. However, due to the difficulties involved in collecting such data, this might necessitate analysis on a case-by-case basis.

Future research on this decision problem can both focus on potential applications of our model in game theoretic and dynamic models and on gathering further insights on the adaptive behavior of terrorist organizations. Prior analyses of learning models in repeated games have shown that agents displaying such behavior can be put into unfavorable positions by their counterparties (Beggs,

2005). Such analyses in the specific area of counter-terrorism thus have the possibility to render informative prescriptive advice. Our results further suggest additional behavioral motives which can be studied. In particular our preliminary results on vicarious learning offer promising avenues of further inquiry.

Concluding, we hope that this paper opens the avenue for more evidence-based counter-terrorism analysis, which is based on real-world data and on more realistic decision processes. This type of analysis may increase the power of decision analysis in providing decision support and risk management, as well as help policy makers in better dealing with the ever-increasing threat imposed by terrorism.

References

Alpern, S., A. Morton, and K. Papadaki (2011). Patrolling games. *Operations Research 59*(5), 1246–1257.

Arin, K. P., O. Lorz, O. F. Reich, and N. Spagnolo (2011). Exploring the dynamics between terrorism and anti-terror spending: Theory and UK-evidence. *Journal of Economic Behavior & Organization 77*(2), 189–202.

Asal, V. and R. K. Rethemeyer (2008). The nature of the beast: Terrorist organizational characteristics and organizational lethality. *Journal of Politics 70*(2), 437–449.

Atran, S. (2003). Genesis of suicide terrorism. *Science 299*(5612), 1534–1539.

Baucells, M., M. Weber, and F. Welfens (2011). Reference-point formation and updating. *Management Science 57*(3), 506–519.

Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of Economic Theory 122*(1), 1–36.

Berman, E. and D. D. Laitin (2008). Religion, terrorism and public goods: Testing the club model. *Journal of Public Economics 92*(10), 1942–1967.

Bhashyam, S. and G. Montibeller (2012). Modeling state-dependent priorities of malicious agents. *Decision Analysis 9*(2), 172–185.

Bhashyam, S. and G. Montibeller (2015). In the opponent's shoes: Increasing the behavioral validity of attackers' judgments in counterterrorism models. *Risk analysis 36*(4), 666–680.

Bier, V. M. and K. Hausken (2011). Endogenizing the sticks and carrots: modeling possible perverse effects of counterterrorism measures. *Annals of Operations Research 186*(1), 39–59.

Bohorquez, J. C., S. Gourley, A. R. Dixon, M. Spagat, and N. F. Johnson (2009). Common ecology quantifies human insurgency. *Nature 462*(7275), 911–914.

Börgers, T. and R. Sarin (2000). Naive reinforcement learning with endogenous aspirations. *International Economic Review 41*(4), 921–950.

Camerer, C. and T.-H. Ho (1999). Experience-weighted attraction learning in normal form games. *Econometrica 67*(4), 827–874.

Caplan, B. (2006). Terrorism: The relevance of the rational choice model. *Public Choice 128*(1-2), 91–107.

Cheung, Y.-W. and D. Friedman (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior 19*(1), 46–76.

Cross, J. G. (1973). A stochastic learning model of economic behavior. *The Quarterly Journal of Economics 87*(2), 239–266.

Cyert, R. and J. March (1963). *A behavioral theory of the firm*. Englewood Cliffs, United States of America: Prentice-Hall.

Denrell, J. (2008). Organizational risk taking: adaptation versus variable risk preferences. *Industrial & Corporate Change 17*(3), 427–466.

Enders, W. and P. Jindapon (2010). Network externalities and the structure of terror networks. *Journal of Conflict Resolution 54*(2), 262–280.

Erev, I. and A. E. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review 88*(4), 848–881.

Feinstein, J. S. and E. H. Kaplan (2010). Analysis of a strategic terror organization. *Journal of Conflict Resolution 54*(2), 281–302.

Franco, L. A. and Hämäläinen (2016). Behavioural operational research: returning to the roots of the or profession. *European Journal of Operational Research 249*(3), 791–795.

Friedman, M. and L. J. Savage (1948). The utility analysis of choices involving risk. *Journal of Political Economy 56*(4), 279–304.

Gill, D. and V. Prowse (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy 124*(6), 1619–1676.

Glaeser, E. L. (2005). The political economy of hatred. *The Quarterly Journal of Economics 120*(1), 45–86.

Greve, H. R. (1998). Performance, aspirations and risky organizational change. *Administrative Science Quarterly 43*(1), 58–86.

Hausken, K. (2018). A cost–benefit analysis of terrorist attacks. *Defence and peace economics 29*(2), 111–129.

Hausken, K. and J. F. Moxnes (2001). Behaviorist stochastic modeling of instrumental learning. *Behavioural processes 56*(2), 121–129.

Herriott, S. R., D. Levinthal, and J. G. March (1985). Learning from experience in organizations. *The American Economic Review 75*(2), 298–302.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science 15*(8), 534–539.

Hoffman, B. (2006). *Inside terrorism*. New York, Unites States of America: Columbia University Press.

Jaspersen, J. and R. Peter (2017). Experiential learning, competitive selection and downside risk: A new perspective on managerial risk-taking. *Organization Science 28*(5), 915–930.

Johnson, N., S. Carran, J. Botner, K. Fontaine, N. Laxague, P. Nuetzel, J. Turnley, and B. Tivnan (2011). Pattern in escalations in insurgent and terrorist activity. *Science 333*(6038), 81–84.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Kaplan, E. H. (2010). Terror queues. *Operations Research 58*(4), 773–784.

Kaplan, E. H. (2012). OR forum—intelligence operations research: the 2010 philip mccord morse lecture. *Operations Research 60*(6), 1297–1309.

Katsikopoulos, K. V. (2014). Bounded rationality: the two cultures. *Journal of Economic Methodology 21*(4), 361–374.

Keller, N. and K. V. Katsikopoulos (2016). On the role of psychological heuristics in operational research; and a demonstration in military stability operations. *European Journal of Operational Research 249*(3), 1063–1073.

Kennedy, A. F. and K. M. Hauksson (2012). *Global Search Engine Marketing*. Que Publishing.

Konrad, K. A. (2004). The investment problem in terrorism. *Economica 71*(283), 449–459.

Köszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Kress, M. and R. Szechtman (2009). Why defeating insurgencies is hard: The effect of intelligence in counterinsurgency operations – a best-case scenario. *Operations Research 57*(3), 578–585.

Lant, T. K. (1992). Aspiration level adaptation: An empirical exploration. *Management Science 38*(5), 623–644.

Luce, R. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Miller, K. D. and W.-R. Chen (2004). Variable organizational risk preferences: Tests of the march-shapira model. *Academy of Management Journal 47*(1), 105–115.

Mookherjee, D. and B. Sopher (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior 19*(1), 97–132.

Neumann, P. (2016). *Radicalized: New jihadists and the threat to the west*. I.B.Tauris.

Newell, A. and P. S. Rosenbloom (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*, pp. 1–55. Lawrence Erlbaum Associates.

Pita, J., M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus (2008). Deployed armor protection: The application of a game theoretic model for security at the los angeles international airport. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems: Industrial track*, pp. 125–132.

Popescu, I. and Y. Wu (2007). Dynamic pricing strategies with reference effects. *Operations Research 55*(3), 413–429.

Powell, R. (2007). Allocating defensive resources with private information about vulnerability. *American Political Science Review 101*(4), 799–809.

Richardson, L. (2006). *What terrorists want*. New York, United States of America: Random House.

Roth, A. E. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior 8*(1), 164–212.

Santifort, C., T. Sandler, and P. T. Brandt (2013). Terrorist attack and target diversity change-points and their drivers. *Journal of Peace Research 50*(1), 75–90.

Seidl, A., E. H. Kaplan, J. P. Caulkins, S. Wrzaczek, and G. Feichtinger (2016). Optimal control of a terror queue. *European Journal of Operational Research 248*(1), 246–256.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics 1*(1), 43–62.

Siebert, J., D. von Winterfeldt, and R. S. John (2016). Identifying and structuring the objectives of the islamic state of iraq and the levant (ISIL) and its followers. *Decision Analysis 13*(1), 26–50.

Sprenger, C. (2016). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy 123*(6), 1456–1499.

START (2015). Global terrorism database. Data file, National Consortium for the Study of Terrorism and Responses to Terrorism.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty 32*(2), 101–130.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplements 2*(4), 1–109.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology 39*(1), 212–222.

Wu, S., Q. Liu, and R. Q. Zhang (2015). The reference effects on a retailer's dynamic pricing and inventory strategies with strategic consumers. *Operations Research 63*(6), 1320–1335.

Zhuang, J. and V. M. Bier (2007). Balancing terrorism and natural disasters-defensive strategy with endogenous attacker effort. *Operations Research 55*(5), 976–991.