Laboratory Experiments on Conflict

Klaus Abbink[1]

1. Introduction

The empirical analysis of conflict has progressed greatly in the last decade (see other chapters in this volume). Despite these successes, there are natural limits to data availability in the real world. Wars are, fortunately, still rare events, and each one has its own history, political background, ethnic composition, etc. Never are there two wars that differ only in one variable, such that the effect of this factor could be isolated. It becomes even more difficult if individual behaviour in conflict is the focus of study. Micro-level data are difficult to come by, not least because it is hard to conduct surveys in a war zone.

In recent years researchers have introduced economic experiments as a complementary approach to observe behaviour in conflict situations. In laboratory experiments it is possible to create analogous – although stylised – environments that mimic real-life conflict scenarios and to obtain data in a controlled manner. This chapter surveys that research.

In a laboratory experiment subjects – typically students as they are readily available on university campuses – make decisions according to well-specified rules that they are given from the experimenter. Depending on the decisions they make, they receive payoffs that are an integral part of the rules of the game. In economic experiments it is common practice to reward subjects in proportion to the payoffs they have achieved in the experiment. This feature ensures that subjects have proper incentives to maximize their payoffs and to make careful decisions. Thus, unlike in most survey studies, subjects play for real money.

The experimental method can be applied for three purposes. The first is to test theoretical models. When modelling a strategic real-life environment, a theorist relies on behavioural assumptions, typically the assumption of fully rational selfish utility (or profit) maximisation. If these assumptions are not met, the theoretical results may be distorted. In the laboratory a rigorous test of the behavioural underpinnings of the model can be carried out. Second, laboratory experiments can substitute field data that often are unavailable when studying conflict. Finally, even if some data are available, laboratory data can be gathered in parallel with field data. The methods are strongly complementary because one method's weaknesses are the other one's strengths. On the one hand, field data are realistic because they are gathered in real life, but they suffer from noise, identification problems, and lack of control. The laboratory, on the other hand, allows the use of a controlled environment in which variations in individual factors can be tested while keeping all others constant. Endogeneity problems do not arise. However, the data are gathered in an artificial environment, which may weaken the external validity of the results.

This article surveys the existing experimental literature on conflict. Conflict experiments are a relatively new field of study, with the first studies carried out in the late 1990s. Overall there are about fifteen experimental economics studies on conflict. The upside of this sparseness is that it is

[1] CBESS, University of East Anglia, Norwich NR4 7TJ, United Kingdom. k.abbink@uea.ac.uk.

still possible to write a survey article that is complete, to the best of my knowledge, and that provides an outline of every completed study. However, this may not be so for too long; several studies are ongoing and were incomplete when this article went to press.

The small number of conflict experiments might also be related to the criterion I have applied to classify a study as a conflict experiment. It is very simple: An experiment is a conflict experiment if the author says so. If the experiment is presented as a model of conflict and title or abstract mention it, it is included; if not, not. This is not to say that experiments that have not been designed as conflict experiments are entirely irrelevant to the field. On the contrary: Because most fundamental games used in experimental research are not specific to one situation they often can be applied to conflict as well. Every public good game can be interpreted as the dilemma one faces when mobilising forces for war. Every bargaining model has something to say about post-conflict negotiations. Every risky choice task can be applied to the choice between a peaceful settlement and the risk of launching an attack. Hence widening the criterion would have either resulted in a general overview of experimental research or a completely arbitrary selection of general purpose models. What this survey should be, however, is a display of how experimental researchers interested in conflict approach the issue.

In some sense, I start the survey with an exception from this rule. One general purpose model, the rent-seeking contest a la Tullock (1980) seems indeed so fundamental to conflict research – and an increasingly popular toy for experimental economists as well –, that it deserves a separate introduction even though it was not originally meant as a model of war. After this prologue I then present the currently available studies on conflict according to four broad themes. Section 3 portrays the tradition of experimental studies on anarchy, based on rent-seeking style models. Section 4 deals with the coordination problem in overthrowing an autocratic ruler. Section 5 is the residual category surveying studies that tackle specific aspects of conflict, without being very much related to one another. Section 6, finally, presents two studies that use experiments in a different way – not to model conflict situations but to gauge the effect that real conflict has on human behaviour. A brief outlook concludes.

2. Prologue: The rent-seeking model

Though they have not been designed to model conflict, rent-seeking style games feature prominently in the experimental (and theoretical) literature on conflict. Introduced by Tullock (1980) a rent-seeking contest involves an exogenous prize *P*, and *n* parties competing to win it. The prize could be a monopoly rent from a government contract (hence the term rent-seeking), victory in a sports competition, but also a resource warring parties fight to secure. At the beginning each competitor decides how much (effort, money, or other resources) to invest in winning the prize. Investment is wasteful, i.e. it creates no wealth as such and is lost should the prize not be won.

At a second stage a lottery is played out that determines the winner of the contest. The probability of winning depends on the relative investments. In the simplest form winning probabilities are proportional to the investments. A more general version adds a decisiveness parameter, which then leads to the famous contest success function

$$p_i = \frac{x_i^m}{\sum_{j=1}^{n} x_j^m}$$

where p is the probability of winning, x the investment, and the exponent m is the decisiveness parameter. It is easy to see how m influences the winning probabilities. If m is large, then investing slightly more than the rivals increases the winning probability drastically. If m is small, then only a much higher investment will secure a high probability of winning. In the extremes, at m=0, the "contest" becomes a pure random draw, and at m=∞, the contest turns into an all-pay auction, in which the highest bidder wins the object but all pay their bids.

Competitors maximise their expected payoff, which is p_iP-x_i. The equilibrium investment for each competitor can be derived as

$$x_i^* = \frac{n-1}{n} mP$$

Thus, a lower decisiveness parameter induces less fighting effort. This is intuitive as the random influence becomes dominant and own effort pays less. In fact, when m=0 investment becomes pointless as the outcome is solely random. Conversely, when m is higher a small advantage in investments induces a huge leap in winning probability, hence the contest becomes more competitive. Note, however, that from a certain point on the investment derived from the formula exceeds the expected value of the prize, such that a rational investor would not take part (or invest 0). Equilibria then exist only in mixed strategies.

While rent-seeking models have mainly been interpreted in economic contexts like lobbying, they suggest themselves as natural workhorses for the study of war. The model captures some essential features of a military conflict. Investment in arms is in itself unproductive, but it increases the chances of winning the war. Despite this, the outcome of the war is still highly uncertain; the strongest army still faces the possibility of losing. So it might be no surprise that the paradigmatic models of military conflict (Hirshleifer (1989, 1991), Skaperdas (1992), Grossman and Kim (1994)) have a contest similar to the rent-seeking game at their core. To put it (perhaps over-)simplistically, the main difference between a typical rent-seeking game and a typical conflict game is that the prize is endogenised, no longer fixed and given. Two (or more) tribes (countries, clans, movements) are endowed with a certain wealth W_i and decide how much they invest in arms. They then fight, and the winner grabs part of the entire wealth of both tribes that has not been invested in arms, i.e. W_1+W_2-x_1-x_2. This means that investment in arms is wasteful in two ways. Like in the rent-seeking games it is an expenditure paid for no consumption value, but beyond that it also shrinks the prize for the winner.

Experimental studies on rent-seeking models are mainly concerned with contest design. Millner and Pratt (1989) and Potters et al. (1998) vary the decisiveness parameter and find the results qualitatively in line with the predictions. Weimann et al (2000), Cadigan (2007), and Sheremeta (2010) study sequential rent-seeking contests. Anderson and Stafford (2003) allow entry and thus endogenise the number of participants. Schmitt et al. (2004) look at a dynamic rent-seeking experiment. Sheremeta and Zhang (2009) and Abbink et al. (2010) examine rent-seeking between teams. A regular finding, somewhat worryingly from the perspective of conflict, is that subjects substantially overinvest, i.e. they behave far more aggressively that they should in equilibrium. It seems that there is more to winning a fight than just the prize itself (see Herrmann and Orzen (2008) for an explanation in terms of spiteful preferences).

3. Anarchy and Conflict

The conflict models derived from rent-seeking games describe a situation in which two parties invest in production and arms, and then fight for the appropriation of the opposition's production. So they can be seen as models of anarchy, in which there is no central state guaranteeing property rights, and the only way to protect own property is by means of force. This is the common theme of the first strand of experimental papers on conflict.

3.1. The Paradox of Power

The first study to be called an experiment on conflict, Durham et al. (1998), is an experimental test of the Paradox of Power (POP, Hirshleifer (1991)). This paradox describes the counterintuitive observation that, although the richer half of the population should have more power than the poorer half, and thus be able to exploit the poor to become even richer, real-life economic policies typically involve redistribution from the rich to the poor. So why do the rich become exploitable? A suggested explanation, derived in a formal conflict model, is that the poor have stronger incentives to fight for redistribution, while for the more productive rich it pays to invest in productive efforts, even if this means that some of the yields will be appropriated.

The experiment is based on the conflict model by Hirshleifer (1989, 1991). Two players, representing the leaders of societal groups, can decide to divide their endowments R_i between a productive effort E_i, and fighting, F_i. After they have done so, total production of both players is shared according to the contest success function $F_i^m/(F_1^m+F_2^m)$. The exponent m is the decisiveness parameter, as in the standard rent-seeking formula.[2]

Depending on the initial allocations R and the decisiveness parameter m the Nash equilibrium prediction involves a strong POP (ex-post wealth is equalised between the groups), a weak POP (the poor can improve their relative position, but remain the poorer group), or no POP (the rich stay rich or even become richer). In an experiment the authors test the theoretical predictions. In no fewer than 12 treatments they vary the endowment (from very unequal to equal), the decisiveness parameter (m=1 or m=4) and the matching protocol (partners and strangers matching).

As a general tendency, fighting efforts in the experiment tend to be lower than predicted by game theory. This is in contrast to the general finding of overinvestment in rent-seeking experiments. However, investments are still close to the equilibrium, and far away from the cooperative solution, which involves minimal investment in the wasteful fighting activity. Fighting is, as predicted, more intense when the decisiveness parameter is high. With regard to POP, the predictions are broadly confirmed. Qualitatively, POP holds when it should hold, and does not when it should not, but the poor generally fail to devote all their effort to fighting when theory would demand that. However, the authors note that in this case the prediction is a corner solution, such that any deviation, for whatever reason, could only bring fighting efforts down. Considering this, the theory does seem to do quite well.

3.2. Choosing to be the predator

[2] Note that in this model production is shared according to the investments. In most rent-seeking models a sole winner is determined by a lottery (though other models exist). For the game theoretic predictions the two variants are the same, as long as players are risk-neutral and maximise expected earnings.

Carter and Anderton (2001) conduct an experiment similar to the seminal paper by Durham et al., but with an asymmetric assignment of roles. Based on the theoretical predator-prey model by Grossman and Kim (1995) they construct an experimental game in which two players invest in production and arms, but only one player's production is subject to seizure. This player, called defence, invests in fortification f in order to protect her own wealth, while the other player, called offence, invests in attacking capacity a. Defence's share of production is $f/(f+\theta a)$, where θ is an attack effectiveness parameter. This parameter θ is the treatment variable in the experiment. Another difference to the model used in Durham et al. is that both players move sequentially, defence first.

Game theoretic analysis predicts that at low values of θ there is no predation; offence invests zero in attack. When θ exceeds a critical value, the solution flips from zero to full predation, i.e. offence invests all its endowment in appropriation and nothing in production. As θ increases further, at some point an area of partial predation is reached, where both players invest in both production and arms. For the three treatments Carter and Anderton choose three values for θ, one from each part of the spectrum.

The results of the experiment provide strong support for the theoretical prediction. In total, more than half of the observations are precisely the theoretical outcome, in the final periods (the game was repeated in fixed pairs eight times) the figure is more than two thirds. Offence chooses the best response given defence's behaviour in 86.3% of cases. Note that offence moves second, knowing defence's choice. Even subjects playing defence manage to respond optimally to offence's choice 64.8% of times, though they have the harder task of having to guess what offence will do. Cooperative off-equilibrium solutions with lower arms investments are rare, but neither do we observe systematic overinvestment as typically seen in rent-seeking experiments.

3.3. Anarchy in the laboratory

In the third study in the tradition of experimental tests of conflict models Duffy and Kim (2005) challenge the theoretical model by Grossman and Kim (2000) and Grossman (2002). Like in Carter and Anderton (2001) the model consists of players who are either producers or predators but not both. Producers have to decide how much to invest in defence, predators only fight and do not produce. The main differences to the previous models are that (1) the model involves larger groups of ten players instead of pairs, and (2) players choose which role to take on, thus the number of producers and predators is endogenous. In equilibrium there will be a certain number of predators, but since all players are identical theory does not predict who will take up which role. This creates a coordination problem for experimental subjects.

Another major innovation in this study is the introduction of a government agent in a second treatment. The government can choose a certain amount that producers have to spend on defence, i.e. that decision is taken from the players. Players then have to choose whether they want to be producers (and spend the government-imposed amount on defence) or predators. The government is incentivised to maximise production. This implies that in equilibrium it will choose the optimal defence level as the one in which defence expenditure is just enough to deter predators, and no predation occurs. In equilibrium outcomes are far more efficient with the government agent than in the anarchic equilibrium without.

The data analysis is a bit hampered by the low number of independent observations (three per treatment) and some heterogeneity across the sessions in the treatment without government. In one of these sessions expenditures on defence were exceptionally low, leading to some "amorphic" outcomes in which there are only predators and nothing is produced. Apart from this observation the data show convergence to the equilibrium prediction, both in the number of predators and the level of defence chosen by producers (or the government). This convergence is quicker and more pronounced in the presence of a government.

3.4. Hobbesian Jungles in real time

The most recent studies on models of anarchy build on the previous efforts, but they do not set out to test the Nash equilibrium prediction of some conflict model. Instead, they design a less structured experimental environment in which decisions are taken in almost continuous time, with user interfaces reminiscent of videogames. Six subjects per session earn an endowment in a quiz. They can decide to invest their endowment in offence or defence or neither. Offence and defence are unproductive, only the endowment that is kept is paid out at the end of the session. A subject can decide to attack another subject, the winner takes the loser's endowment with a probability proportional to the share of the attacker's offensive and the target's defensive investments. While this general setup is similar to previous experiments, an important difference is that in Powell and Wilson (2008) subjects can do everything at any time, not following an imposed move sequence. This unstructured setting should mimic a Hobbesian jungle, a state of anarchy in which no societal order exists and, at least potentially, there is war of all against all at any time. The authors have to sacrifice game theoretic tractability, but as Powell and Wilson write, "replacing the phenomenon of the Hobbesian jungle with an analytically solvable model is [...] perhaps oxymoronic."

In a follow-up study, Smith et al. (2009) allow subjects to form groups. Such endogenous groups pool their defensive or offensive capabilities. The treatment is designed to test a philosophical theory put forward by Nozick (1974). According to this theory, even in absence of a central order individuals would form protective associations, which would then concentrate and eventually form a monopolistic order from the state of anarchy. Smith et al. also introduce communication between participants in order to facilitate the formation of groups or cooperative, non-violent agreements.

The continuous-time approach is novel and fascinating. It provides an avenue of studying conflict behaviour in an environment that is still controlled, but richer and more natural than the usual approach of very stylised games. The richness of the environment comes at the cost that the results of the experiments are somewhat hard to interpret. The problem is not the lack of a game-theoretic benchmark (the importance of which is perhaps overrated in the experimental economics community), but the enormous heterogeneity of the outcomes. In the original data efficiency, measured as the fraction of potential earnings actually realised, ranges from dismal (14%) to rather high (71%). Some sessions learn over time to cooperate, in others efficiency deteriorates. The introduction of groups does not significantly affect efficiency, but with few data points and strong heterogeneity this might be due to lack of statistical power. The overall impression is that anarchic environments neither lead to a constant war of all against all nor to a utopia of universal cooperation. The difficult results should not be seen as discouraging. Experiments with simple games also often show heterogeneous results, but since there is a huge body of existing data they seem

easier to understand. Such a reference does not exist in the present case, such that more data are needed for more systematic inference.

4. The Coordination Problem in Conflict

The studies in the second group of conflict experiments depart from the model of anarchy towards the opposite manifestation of power: dictatorship. Once an autocratic ruler has come into power, how does he stay there? What keeps the oppressed and exploited masses from rising up and throw him out?

Several scholars point at a coordination problem that citizens and group leaders face. For the individual it is not advisable to speak up or fight against an authoritarian regime as long as he is alone. Only if sufficiently many others join forces, a challenge can be successful. So it can happen that even though everybody wants revolt, and together people would be able to, nothing happens because no-one expects others to turn to the streets. This coordination problem lies at the heart of the second strand of experiments on conflict.

4.1. Lab revolutions

A hallmark of civil uprisings is their unpredictability. The noughties saw a string of successful revolutions (Yugoslavia 2001, Georgia 2003, Ukraine 2004, Kyrgyzstan 2005 and 2010, Tunisia, Egypt 2011), as well as some unsuccessful uprisings (Burma 2007, Iran 2009). Arguably, all these events emerged more or less out of the blue, raising the question: Why did it happen, and why did it happen now? Empirical research into this question is hampered by the rarity of the events, many idiosyncratic circumstances and the unobservability of those revolts that did *not* happen, though circumstances may have been right. Abbink and Pezzini (2005) point at the game theoretic structure of civil uprisings to explain their unpredictability. For a revolution to be successful it not only needs a common sentiment that action is desirable, but also the simultaneous belief that a critical number of people will take the street at the same time. When and how beliefs flip towards a spontaneous uprising is not foreseeable, but it is possible to empirically determine which factors make a revolt more likely.

Abbink and Pezzini conduct a combined analysis of survey data and a controlled laboratory experiment to find out the relationship between revolutionary *preferences* and *actions*. To assess preferences an econometric analysis of data from the World Value Survey is performed, in particular questions whether the subject thinks revolutionary changes in politics are desirable. In order to study revolutionary actions the authors design a three stage game between one governor and seven citizens. At the first stage the governor decides about three policy dimensions: A distribution of income, the restriction of freedom of communication, and the level of repression against unsuccessful opposition. At the second stage citizens can send messages about whether or not they wish to oppose the governor. If the governor has chosen a restrictive communication policy, then citizens expressing opposition face the risk of being fined. At a third stage citizens then decide of whether they actually oppose the policy. If fewer than five citizens oppose, then the governor stays in power, and unsuccessful rebels receive a fine. If five or more citizens oppose, then the citizens who didn't are worse off (as they would be accused of being collaborators with the old regime), and the governor is replaced by a randomly selected new one.

The results show that at the level of actions greedy governors are very likely to be overthrown, while restricting communication and punishing opposition dampens the citizens' willingness to revolt. At the level of preferences an analogous result is obtained for the ruler's greed, which tends to fire up revolutionary sentiments. For freedom of communication and repression, however, restrictive policies have opposite effects on preferences and actions. While survey data show that they make citizens more willing to revolt, the experimental results show that they lead to a lower likelihood of these actions actually being carried out. Thus, dictators who want to stay in power should not steal too much, but be tough against opposition.

4.2. Riots

In 2005, immigrant youths in the *banlieues*, the deprived satellite towns of French cities which are now inhabited by low-income groups, rioted for several weeks, burning cars and destroying public buildings. Many commentators saw inequality and lack of perspective as the root cause of the uprising. Inspired by these events, Abbink, Masclet, and Mirza (2006) study the relationship between inequality and riots in a controlled laboratory experiment. They devise a two-stage game between two groups of players, an advantaged group and a disadvantaged one. At the first stage of the game players earn an income in a rent-seeking game. The disadvantaged players have a much lower productivity of their effort in the contest, in fact the game is parameterised in a way that they should not invest anything at all and the whole income goes to the advantaged players (in this experiment, behaviour in the rent-seeking contest as such is not of interest).

At the second stage members of the groups can decide to riot or not. Similar to Abbink and Pezzini's setting, rioting is a multiplayer coordination game. A riot is only successful if many people show up (in which case it is a best response to join), if too few show up the rioters get punished. In contrast to the previous study rioting is purely destructive; the only effect of a successful riot is that the other group's income is reduced. Thus the experimental design rules out self-interested goals like gaining power or even public attention.

In three treatments the authors vary the degree of inequality, including a scenario without any inequality. As expected, when disadvantaged groups clash with advantaged ones, it is the former who are more likely to riot. However, the comparison across treatments yields an unexpected result: More inequality induces *less* rioting. In fact, riots are most likely when two equally advantaged groups clash. The authors offer two competing explanations for this result. First, disadvantaged groups may refrain from rioting out of fear of a counterattack in later rounds (the experiment was repeated with fixed groups). Alternatively, disadvantaged players might accommodate with their unfortunate position and resign. To test these hypotheses against one another, the authors conduct a new treatment with strangers matching, i.e. the groups are re-shuffled every round. Fear of counterattacks does not apply here, nevertheless riot frequencies remain virtually unchanged. Hence, resignation can be seen as the most likely explanation for the phenomenon.

4.3. Divide and Conquer in the laboratory

"Divide and Conquer" (DAC) describes the age-old strategy of military leaders to gather support from some leaders of rivalling groups in order to steal from the others. Weingast (1997) has devised a formal game-theoretic model of this policy, exposing the coordination problem underlying it. In his game a ruler faces two leaders of groups, e.g. ethnicities or classes. At a first stage the ruler decides

to either transgress against the groups (which means looting their resources) or not to do so. If he transgresses, he can also choose to transgress against both groups or only one of them. If he transgresses against one, then he can make the other group slightly better off in order to win its support.

At a second stage the group leaders decide simultaneously whether to acquiesce or to challenge the ruler. The crucial point is that a challenge is only successful (meaning that the ruler is punished and the groups keep their resources) if both leaders challenge. If only one group leader challenges then the challenging leader incurs a cost, since the revolt is unsuccessful and he will be punished.

Weingast shows that in all equilibria the ruler transgresses, because the group leader who is not robbed has a dominant strategy to acquiesce. Subgame perfect equilibria always involve transgressions against one at least group (it can be both groups, but this is a less plausible equilibrium). The equilibria in which the ruler transgresses against one of the groups, who then do not challenge him, are the DAC equilibria of the game.

Using Weingast's game Cason and Mui (2007, 2009) test whether experimental subjects can keep rulers from transgressing. Although this is not an equilibrium in the one-shot game, there is good reason to believe that other-regarding preferences and group solidarity, for which there is ample evidence in the experimental literature, might be strong enough forces to let group leaders coordinate on the off-equilibrium challenge, and the threat might deter rulers from transgressing. The authors analyse the effect of two institutions: pre-play communication and repeated play. Pre-play communication consists of a single standardised message in which leaders could signal their intention to challenge or not. In the 2007 study the focus is on various modes of communication in a random re-matching environment. Either the leaders can exchange private messages between each other, or they can make public announcements, which the ruler receives as well. The 2009 study focuses on repeated play, which is varied in several ways, with a finite known horizon or a probabilistic stopping rule (which mimics infinitely repeated play).

The results show that both communication and repeated play are effective tools to reduce transgressions by the rulers. Interestingly, the most effective communication mode is private ex-ante communication, i.e. before the ruler has made his decision the group leaders can exchange private messages. Thus, although the ruler does not even get the messages they seem to be a more credible threat than public announcements. Without communication outcomes without transgression are rare. Cason and Mui (2007) points out that public communication may be less effective because it enables the ruler to practice divide-and-conquer.

The effect of repeated play as such is much weaker. The most effective repetition, repeated play over a long horizon, reduces transgression rates from 93.5% to 67.4%, while probabilistic repetition (in the experiment the parameters were chosen such that the strings of games were rather short), has hardly any effect. The effects of repeated play and communication seem to be largely additive, i.e. the effect of combining the two is similar to the sum of the two separate effects.

5. How to Prevent Conflict

This section could also be entitled "Miscellaneous topics", since the studies here do not really follow a common thread (besides modelling conflict situations). They represent attempts at tackling

specific environments and problems surrounding conflict, by means of laboratory experiments. A common question is which environments make conflict more likely or more intense, and thus the results may provide some insight into possible measures to prevent conflict.

5.1. Independence struggles

Abbink and Brandts (2006) use a variant of a rent-seeking game to model a specific type of civil conflict, the struggle for autonomy of minorities in multi-ethnic countries. They construct a rather complex multi-player multi-stage game to study conditions under which peaceful solutions do or do not exist. Their game consists of n players representing the majority ethnicity in region A, and a minority of m players in a region B striving for independence. There are different levels of autonomy that can be granted to region B, between full centralisation and full independence. The preferences of the citizens A and B are opposed: The citizens B prefer more autonomy over less, for the citizens B the reverse holds.

In a voting game the citizens of both regions can negotiate an autonomy level acceptable to both regions. First, a citizen A proposes an autonomy level, over which all citizens vote. If the proposal gets a majority, it is the new status quo, otherwise full centralisation prevails. In any case, there is another voting stage in which the citizens B decide whether or not to enter conflict for full independence. If they decide against conflict, then the status quo is implemented, if they choose conflict, then a team rent-seeking game is played. All citizens invest in their respective war chests. The winner of the contest is determined by a random draw, with winning probabilities proportional to the aggregate contributions to the respective war chests. The winning region gets its most preferred outcome, full centralisation if A wins, full autonomy on a victory of region B.

The authors analyse the conditions under which peaceful interior solutions are possible. This depends on the relative preferences for autonomy. If the preferences are concave, i.e. if there is an interior solution that gives both regions a higher payoff than the expected conflict payoff, then a peaceful solution should emerge in the game theoretic equilibrium. If preferences are very convex, then no peaceful solution is possible and conflict is inevitable.

The experimental results refute the theoretical prediction. The authors conduct sessions with 6 citizens A and 3 citizens B who repeatedly play the game. Although the parameterisation should be very favourable to peace, with strongly concave preferences and very salient interior solutions, conflict is very frequent. In the three treatments, which vary the concavity of preferences, conflict frequencies range from 38% to 64%. But not only is conflict much more frequent than predicted, but also much more intense. When the conflict stage is reached, contributions to the war chest are more than four times higher than in the subgame equilibrium.

The over-fighting observed in the experiment suggests a strong emotional charge that the strategic environment loads on players. This may be due to the fact that the game was presented to the subjects in the context of an independence struggle and conducted with Spanish and Catalan students in Barcelona. So the authors conduct additional experiments in Amsterdam, where there are no separatist movements. In both locations, sessions with abstractly worded instructions were also conducted. However, substantial over-fighting remains. Conflict frequency is not significantly affected by either wording or subject pool. Wording reduces conflict intensity in the Barcelona

subject pool, but not in Amsterdam. Hence the real-world context only has an effect if subjects are already emotionally involved in a separatist struggle.

5.2. The shadow of the future

A standard result from the theory of infinitely repeated games is that cooperation becomes easier to establish, and conflict less likely, when players place a high value on future payoffs, i.e. if future payoffs are discounted very little. McBride and Skaperdas argue that this result cannot be immediately transferred to conflict situations. This is because conflict between tribes or countries has a strong dynamic component beyond mere repetition: The winner of a conflict gains relative strength for future encounters, while the loser is weakened. As a result conflict can even become more likely when the "shadow of the future" looms large.

McBride and Skaperdas (2009) test this hypothesis in a very simple two-person game. In each round players choose either peaceful settlement, in which both players receive 20 money units each, or conflict, in which the winner gets 40 and the loser makes a net loss of 30, each occurring with equal probability. Conflict breaks out if at least one player chooses conflict.

The expected payoff of conflict is 5, thus for this one period conflict does not pay. However, the main feature of the design is that the decision in period 1 has implications for future periods. In particular, the winner of the conflict in period 1 gets 100 points in every future round, the loser 0. If there had been peaceful settlement in period 1, then each player receives 20 in every future period. Whether or not there will be a future period is determined by random draw at the end of each period. The continuation probability is the main treatment variable, it takes values of 0, 0.5 and 0.75. It is easy to see that with these payoffs conflict becomes more attractive with higher continuation probabilities. In fact, risk-neutral players would choose conflict for p=0.5 and p=0.75, and settlement when p=0.[3]

The results show that the qualitative predictions are confirmed: Players choose more conflict with rising continuation probabilities. The effect is, however, less sharp than the payoff structure would suggest. With p=0, when only extremely risk-loving players would want conflict, it is observed in a quarter of all periods (27%). With p=0.75 the expected payoff of conflict is almost twice as high as that of peaceful settlement, still not everybody chooses conflict, but about a third do not (conflict frequency 66%). By and large, however, the theoretical predictions are supported by the data.

5.3. An analysis of counterterrorism strategies

Colombier et al (2009) emphasise the public good character of international counterterrorism strategies. They consider two basic strategies countries can adopt. A country can invest in unilateral measures, for example border control efforts. These decrease the probability of being hit by a terrorist attack, but they do not help other countries. In the extreme, as modelled in the study, they even exert a negative externality – border control measures may help to keep terrorists out of the own territory, but with the effect of terrorists striking elsewhere instead. In contrast, investment in international cooperation, e.g. concerted policing or military action, benefits the whole international

[3] Strictly speaking the prediction is indeterminate. Conflict is always an equilibrium outcome, but one in weakly dominated strategies and thus somewhat implausible.

community in lowering the probability of every country of being hit. This generates an incentive to free-ride, since it allows a country to benefit from others' efforts without investing itself.

To keep things simple (and in the tradition of the vast literature of public good experiments) the authors set up a linear model in which the equilibrium prediction is a corner solution. In the model it is a dominant strategy to invest in national protection only, and nothing in international collaboration. The pareto optimum, on the other hand, is achieved when all countries invest everything in international cooperation. This is an extreme representation of the free-rider effect and arguably a departure from reality, where the individually optimal solution is likely to be a combination of national and international efforts. However, this linear setting has two advantages. First, it is much easier to handle in an experiment than a non-linear setting with interior solutions, and second, it models the tension between national egoism and international solidarity in an idealised manner and therefore allows to sharply identify behavioural patterns arising from it.

The experiment consists of three treatments. In the baseline treatment, countries decide how much to spend on either national or international protection. Investment in either means that the probability of suffering a terrorist attack (which is modelled as an exogenous random event, as strategic behaviour of terrorists is not of interest here). The difference is that if a country invests in national security, the probability of other countries suffering an attack increases (thus creating a negative externality); investment in international activities decreases the likelihood of an attack for all countries (the externality is positive). The two other treatments test institutions that may work to foster international collaboration. In the punishment condition, all players are informed how much the other countries have invested, and then have an opportunity to punish other countries (representing economic or political sanctions). In the experiment punishment means that the target's payoff is reduced, at a cost for the punisher (in the spirit of Fehr and Gächter (2000)). Every 20 money units reduction for the target cost the punisher 5 money units. The third treatment is analogous, but instead of punishing others players now can reward others, again at a cost. If a player decides to reward another player, the beneficiary's income is increased by 20 for every 5 money units the benefactor invests to reward.

Without punishment or reward opportunities subjects do not manage to achieve much international collaboration in the experiment. Most subjects invest only in inefficient national security; few very risk-seeking players do not invest in security at all. Both punishment and reward improve international collaboration substantially, where rewards are much more effective than sanctions. The latter effect is certainly helped by the efficiency gain generated through rewards: For each money unit paid for rewards the beneficiary receives four. This means three money units fall from heaven whenever a player decides to reward. This setting retains some symmetry with the punishment treatment, but in reality it is not easy to see how a reward system with such huge efficiency gains can be implemented.

5.4. Should the winner be greedy or generous?

After the fighting has ended, trouble is not always over. Sometimes, as the recent case of Iraq shows, managing the post-conflict stage can even be more problematic than winning the war itself. Should the winner be tough and exploit as much as possible from the victim, or be generous and even help. Can the loser eventually win a lost war? The study by Lacomba et al. (2008) addresses post-conflict behaviour. The study combines a conflict model as in Hirshleifer (1988, 1991) and

Skaperdas (1992) with a second stage in which players have additional choices after they have learned whether they have won or lost the war.

As in previous models, the first stage of the game involves a lottery contest. Two countries divide their endowment between investment in a productive activity and investment in arms. After these decisions have been made simultaneously, a lottery is played out that the respective countries win with probabilities proportional to their investments in arms. The prize fought over is the production of both countries. If both countries invest nothing, then there is no war. Thus, war is avoidable, but this is a precarious state of the world: If one country invests nothing, then the other country can win the war with certainty by investing minimally in arms – hence getting the entire pie virtually for free.

The novelty of this setting lies in the additional stages the authors introduce. While in the standard conflict model described above the winner automatically gets the whole production, Lacomba et al. make this outcome an endogenous choice. The winners get the right to appropriate as much as they want, but they can choose the entire amount. Further, depending on the treatment the loser can still influence how much the winner gets by destroying part or all of her production. The authors consider three variants of the game. In the *Complete Surrender* (CS) treatment the winner unilaterally decides how much to take. Obviously, a self-interested winner would not take less than the entire amount, but the authors hypothesise that winners may leave some money to signal their goodwill or their generosity.

In the other two treatments the loser makes a decision as well. The *Resistance* (RE) treatment adds to the CS treatment a third stage in which the loser, after having learned how much the winner has taken, can destroy part or all of the total production. This reflects a situation in which the winner, while appropriating the loser's wealth, still cannot completely control the loser's production and the loser can resort to sabotage or make life as difficult as possible for the winner, even if the loser has no immediate gain from it. If the winner has left some part then destruction is costly to the loser. Still, losers may choose to destroy in order to punish the winner or to enforce more generous behaviour in the future.

The third treatment, called *Scorched Earth* (SE) is analogous to the RE treatment, but with the second and third stage of the game reversed. Here, the loser decides first how much to destroy, and afterwards, knowing the loser's destruction, the winner decides how much to take from what is left. The payoff consequences are the same as in RE, but the two settings may well be behaviourally different, since the loser must decide before knowing how generous or greedy the winner will be.

The games are played repeatedly for 10 times, under either partners (the same two subjects meeting again and again) or strangers matching (pairs are randomly reshuffled every round). In the baseline CS treatment, without destruction, winners almost always take 100% of the pie. This is what self-interested players would do, but the result is nevertheless surprising. The last stage of the CS treatment is basically a dictator game as it is extensively studied in experimental research: One player splits a given pie between Self and an anonymous recipient. In those games it is common that players give away substantial amounts against their own self-interest, most likely out of altruistic motives or fairness considerations. Apparently, having just fought a war against each other suppresses such motives quite effectively.

An interesting comparison is the one between take/destruction rates in the SE and the RE treatments. Recall that these treatments are the same in actions and payoff consequences, just with a reversed order of moves. Here the SE treatment, in which the loser decides first, triggers higher destruction rates and higher take rates than the RE treatment in which the winner moves first. Take rates in SE are almost as high as in the baseline CS treatment where the loser has no power at all. It seems that if the winner moves first, he can show generosity by leaving money, which the loser can reciprocate by not destroying. If the loser has to make the negative move first, this creates an atmosphere of mistrust.

Over time, the differences in post-conflict behaviour have an effect on conflict intensity in the first stage of the game. Investments in arms are highest in the CS treatment, lowest in the RE treatment and intermediate in SE. Hence an environment that treats losers better is less conducive to intense wars than one in which winners can take all.

6. Post-Conflict Behaviour

In the final section of this overview I present two studies that address the issue of conflict with laboratory experiment, but from a completely different angle. They do not attempt at modelling conflict as an experimental game, but rather use standard experimental games as measurement tools to gauge the impact of conflict on human behaviour. These games are standard tools in experimental research, like dictator games, trust games, or public good games.

6.1. Can fairness recover from a bloody war?

Whitt and Wilson (2007) study whether civil relationships can be sustained, or rejuvenated, between ethnic groups that previously were at war against each other. The case they study is post-war Bosnia. Between 1992 and 1995, ethnic hostilities between Bosnjaks, Serbs and Croats had claimed more than 100,000 casualties and displaced millions (exact figures are disputed). Seven years after the end of the war, Whitt and Wilson conducted an experiment designed to study the interethnic relationships between the three groups. The authors are particularly interested in fairness norms, which they test with the tried and tested instrument of the dictator game. In this “game” (strictly speaking an individual decision without strategic interaction), a sender is asked to divide an amount of money between Self and a second person, the recipient. The recipient is passive, s/he simply gets whatever the sender allocated to them. Selfishly rational senders would keep the entire pie to themselves, however, in a plethora of experiments (see Camerer (2003) for an overview) it has been found that many senders give substantial amounts, often up to the equal split. The game is therefore frequently used as a workhorse to measure norms of fairness and altruistic motives.

Whitt and Wilson vary the ethnicity of senders and recipients in a within-subject design. Senders make two dictator decisions. First they decide how much to allocate to an anonymous recipient of their own ethnicity. After that, they make a second analogous decision, but this time they are truthfully told from which of the other two ethnic groups the anonymous recipient is drawn. The difference between the amount sent to a member of their own and that sent to a member of a different ethnic group measures the extent of out-group discrimination prevalent within an ethnic group.

The results show that subjects give substantial amounts to the recipients, even if the recipients come from a formerly hostile ethnic group. The amounts are comparable to standard dictator games played with other, mainly student subjects in the literature. Hence it appears that seven years after the war, fairness norms have been either sustained or re-established. This is the encouraging side of the data. However, the data also reveal substantial discrimination against other ethnicities. All three groups of subjects give significantly less to members of the other groups than to co-ethnics. Further, Croats tend to give less than Bosnjaks or Serbs, both to co-ethnics and others.

The result that fairness norms and altruistic motives are intact seven years after the end of the war is encouraging. However, the absence of a proper control condition makes the results hard to interpret. We don't know how the people in Bosnia would have behaved before the war, since naturally, no-one has conducted the experiment back then. So we do not know either to what extent the observed discrimination against members of other ethnicities is an expression of prevailing hostility or a more ordinary in-group identification effect.

6.2. Development and cooperation after the war

The approach these authors take is similar to that of Whitt and Wilson (2007). Coleman and Lopez (2010) also use simple experimental games to assess cooperation in post-conflict societies. In contrast to the previous study they do not set out to study the effect of conflict as such, but the impact of a policy instrument that has been designed for Colombia to improve the chances of lasting peace by "strengthening local democratic institutions, instilling values of human rights, and providing basic economic goods and human services." This programme, called "Peace and Development", funds a variety of activities ranging from supporting farmers to increase their crop yields to providing civic education or arts. The programme has been introduced in 1995 in one region of the country and expanded in 2002, 2004 and 2006.

The study by Coleman and Lopez is designed to assess the impact of the Peace and Development programme. Since most of the policy goals are hard to measure, the authors chose to look at the willingness and ability of the beneficiaries to establish cooperation within the society. To measure cooperation they conduct three simple experimental games. The games are chosen to measure different levels between "insular" and "inclusive" cooperation. The authors acknowledge that cooperation is not always beneficial. If people cooperate only with their very narrow in-group (insular cooperation) at the expense of others it can be harmful to society, like, for example, cooperation within organised gangs. Beneficial cooperation should extend to the entire society (inclusive cooperation).

The three games reflect different levels of inclusiveness. The game that stands for most insular cooperation is the two-player trust game. The first mover, the trustor, receives an endowment that he can either keep or send to the second mover, the trustee. Every money unit sent to the trustee is tripled by the experimenter, creating a surplus. The trustee then decides how much of the money she has received to return to the trustor. The multiplier ensures that both players benefit from trust and reciprocity. However, purely selfish trustees would not send back any money, and in turn trustors would not send anything if they believe that the trustee is selfish.

Following the trust game, subjects are asked whether they wish to contribute their earnings (or part of them) to a public service. This is a real institution like a school or health centre. This contribution

measures the most inclusive type of cooperation, since the benefit is (typically) very widespread to the whole society. A multiplayer public good game covers the middle ground between the two. In this game players are given an endowment, which they can either keep or pay into a common pool. If a player invests, he himself gets less out of the investment than if he had kept the money, but everybody else also benefits. The pareto-optimal solution is realised if everybody invests, but individually the dominant strategy is to keep the endowment. Unlike in standard public goods experiments, where the number of players is typically between three and five, Coleman and Lopez play their experiment with large groups of more than 20 players.

To evaluate the success of the Peace and Development programme a control group is needed. So the authors conduct their experiment in five regions in which the programme has been in effect, and in another one which has been identified as a target region, but payments have not yet been made. This qualifies the region as a control group, since it fulfils the criteria for the introduction of the programme (hence avoids selection effects), but since the programme has not yet started it cannot have had any effect. While this is not a fully randomised trial (in which treatment and control groups are randomly assigned before the experiment), the presence of the control region constitutes a useful natural experiment to test the effects of the programme on cooperation.

The results show that the Peace and Development programme had indeed a moderate, but significant effect in fostering cooperation among the beneficiaries. Surprisingly, the correlation between the different types of cooperation is only weak. The three games indeed seem to measure different types of cooperation and not a general notion of prosociality.

7. Outlook

Experimental conflict research is very much in its infancy and only just gathering momentum. The number of studies on conflict-related topics is still very small, and the range of topics wide, such that the present state of the literature cannot be more than a collection of snapshots. A unifying theme or a standard paradigm have not yet evolved, with the possible exception of the strand of literature on models of anarchy, which do follow some kind of a common thread.

In the future this is likely to change, as the number of experimental studies grows. I see a great potential for this new method, since experiments allow doing things that couldn't be done otherwise. Mere theoretical reasoning leaves open the question of behavioural validity, field data analysis often lacks control over specific variables, which makes ceteris-paribus analysis difficult. Laboratory experiments can bridge the gap. We can create controlled and replicable environments of conflict, changing one variable while keeping everything else constant.

As always in life, these advantages come at a price. The experimental games are stylised models of war, they are not real war. No-one dies in the experimental laboratory. The question arises whether we can take results from lab wars seriously, given that subjects play for small amounts of money while real wars are matters of life or death. There can be no doubt that this is a valid question. It is eventually the question of external validity of experimental results, the generalisability of lab insights beyond the confines of the laboratory.

While the question is justified I do not think that it should put us off. It is a misunderstanding of the experimental approach to believe that the environment we recreate in the lab must or should be

exactly the one in reality, and if there are differences they immediately invalidate the results. In most cases experimental inference comes from treatment comparisons. When we look at data from a conflict experiment, for example, it is normally meaningless whether conflict frequencies are, say, 30% or 60%. The parameters of the game are invented anyway, and not aimed at reflecting any realistic values. But if we compare two experimental conflict settings and find that conflict frequencies are 30% under one condition and 60% under another, what we can infer is that the latter environment is more prone to conflict than the other. Since the two conditions are exactly the same except for the variable that we have changed, we can also infer that the change in the variable has caused the increased likelihood of conflict. Do we have reason to believe that a similar change of a similar variable in a real-life environment would also induce an increase in the likelihood of conflict? Well, probably we have more reason to believe that than to believe the opposite – that a variable that is shown to be conflict-increasing in the lab will reduce conflict in reality. Of course, we cannot be totally sure. Perhaps the experimental model has left out a crucial feature that turns all effects upside down. Perhaps the experiment was poorly designed and emphasises an effect that is irrelevant in the real world. It is possible to make wrong inferences from experiments, but this is not specific to the method – it is just as possible to be misled by a peculiar theory or an ill-specified econometric model. Thus, I want to argue that despite the differences between lab and reality there are many lessons to be learned from well-designed experiments.

Eventually the question of external validity needs a stronger link between the laboratory and the field. This is of course difficult because often the very reason to conduct experiments is that we cannot observe in the field what we want to study. However, the growing popularity of field experiments (see Harrison and List (2004) for an introduction) shows that there is room to occupy the wide space between the lab and the field. In section 4 of this survey experimenters have taken lab games into the world and gained interesting insights from it. So as the experimental literature on conflict becomes larger and more diverse, there is reason to expect that the experimental method will be a more and more useful tool to understand conflict.

References

Abbink, K, and J Brandts (2009): Political Autonomy and Independence: Theory and Experimental Evidence, Working paper, University of Amsterdam.

Abbink, K, J Brandts, B Herrmann, and H Orzen (2009): Inter-Group Conflict and Intra-Group Punishment in an Experimental Contest Game. *American Economic Review*, 100, 420-447.

Abbink, K, D Masclet, and D Mirza (2008): Inequality and Riots – Experimental Evidence. Working paper, University of East Anglia.

Abbink K and S Pezzini (2005): Determinants of Revolt: Evidence from Survey and Laboratory Data. Working paper, University of Nottingham.

Anderson, LR, and SL Stafford (2003): An experimental analysis of rent seeking under varying competitive conditions. *Public Choice*, 115, 199-216.

Cadigan, J (2007): Two-Stage Team Rent-Seeking: Experimental Analysis. *Southern Economic Journal*, 74, 85-103.

Camerer, C (2003): Behavioral Game Theory: Experiments on Strategic Interaction, Princeton: Princeton University Press.

Carter, JR, and CH Anderton (2001): An experimental test of a predator-prey model of appropriation. *Journal of Economic Behavior and Organization*, 45, 83-97.

Cason, TN, and V-L Mui (2007): Communication and Coordination in the Laboratory Collective Resistance Game. Experimental Economics, 10, 251-267.

Cason, TN, and V-L Mui (2009): Coordinating Collective Resistance through Communication and Repeated Interaction. Working paper, Monash University.

Coleman, EA, and M-C Lopez (2010): Reconstructing Cooperation from Civil Conflict: Experimental Evidence from Colombian Development Policy. Working paper, Florida State University.

Colombier, N, D Masclet, D Mirza, and C Montmarquette (2009): Global Security Policies Against Terrorism and the Free Riding Problem: An Experimental Approach. Working paper, University of Rennes 1.

Duffy, J, and M Kim (2005): Anarchy in the laboratory (and the role of the state). *Journal of Economic Behavior and Organization*, 56, 297-329.

Durham, Y, J Hirshleifer, and VL Smith (1998): Do the Rich Get Richer and the Poor Poorer? Experimental Tests of a Model of Power. *American Economic Review*, 88, 970-983.

Grossman, HI (2002): Make us a king: anarchy, predation, and the state. European Journal of Political Economy, 18, 31-46.

Grossman, HI, and M Kim (1995): Swords or plowshares? A theory of the security of claims to property. Journal of Political Economy, 103, 1275-1288.

Harrison, GW, and JA List (2004). Field experiments. Journal of Economic Literature, 42, 1009-1055.

Herrmann, B, and H Orzen (2008): The appearance of homo rivalis: Social preferences and the nature of rent-seeking. Working paper, University of Nottingham.

Hirshleifer, J (1989): Conflict and rent-seeking success functions: Ratio vs. different models of relative success. *Public Choice*, 63, 101-112.

Hirshleifer, J (1991): The paradox of power. Economics and Politics, 3, 177-200.

Lacomba, JA, FM Lagos, E Reuben, and F van Winden (2008): After the War: An experimental study of post-conflict behaviour. Working paper, University of Amsterdam.

McBride, M, and S Skaperdas (2009): Conflict, Settlement, and the Shadow of the Future. Working paper, University of California Irvine.

Millner, EL, and MD Pratt (1989): An experimental investigation of efficient rent-seeking. *Public Choice*, 62, 139-151.

Nozick, R (1974): Anarchy, State, and Utopia. New York:Basic Books

Potters, J, CG de Vries, and F van Winden (1998): An experimental examination of efficient rent-seeking. *European Journal of Political Economy*, 14, 783-800.

Powell, B, and BJ Wilson (2008): An experimental investigation of Hobbesian jungles. *Journal of Economic Behavior and Organization*, 66, 669-686.

Schmitt, P, R Shupp, K Swope, and J Cadigan (2004): Multi-period rent-seeking contests with carryover: Theory and experimental evidence. *Economics of Governance*, 5, 187-211.

Sheremeta, RM (2010): Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior*, 68, 731-747.

Sheremeta, RM, and J Zhang (2009): Can Groups Solve the Problem of Overbidding in Contests? Working paper, McMaster University.

Skaperdas, S (1992): Cooperation, conflict, and power in the absence of property rights. *American Economic Review*, 82, 720-739.

Smith, AC, DB Skarbek, and BJ Wilson (2009): Anarchy, Groups, and Conflict: An Experiment on the Emergence of Protective Associations. Working paper, George Mason University.

Tullock, G(1980): Efficient rent-seeking. In: JM Buchanan, RD Tollison, and G Tullock (eds): Toward a theory of the rent-seeking society. College Station: Texas A&M Press.

Weimann, J, C-L Yang, and C Vogt (2000): An experiment on sequential rent-seeking. *Journal of Economic Behavior and Organization*, 41, 405-426.

Weingast, B (1997): The political foundations of democracy and the rule of law. American Political Science Review, 91: 245-263.

Whitt, S, and RK Wilson (2007): The Dictator Game, Fairness and Ethnicity in Postwar Bosnia. *American Journal of Political Science*, 51, 655-668.