Terrorism prevention: a general model

Franz Dietrich[1]
University of Konstanz
February 2004

Abstract

The trade-off between capturing terrorists and, by this action, possibly creating new terrorists, may be analysed based on models such as the following one. The probability of a (random) person being a "free terrorist" equals that of a person being a "terrorist" multiplied with that of a "terrorist" being "free"; formally, P("free terrorist") = P("terrorist") × P("free" | "terrorist"). To reduce this product, the policy should reduce both factors. Cause-related policies reduce P("terrorist") by reducing people's inclination towards terrorism, involving measures such as raising the standard of living. Symptom-related policies reduce P("free" | "terrorist") by detaining terrorists. But symptom-related policies also affect P("terrorist"), through (desirable) deterrence and (undesirable) "hate effects" on people's inclination towards terrorism. If "hate effects" dominate over deterrence, more toughness overall increases P("terrorist"), possibly overcompensating the reduction of P("free" | "terrorist"). This and other models point towards a new method for analysing terrorism policy choice.

The political debate about terrorism prevention is dominated by two seemingly opposed worries: "How can we prevent people from becoming terrorists?", and "How can we neutralise the threat coming from existing terrorists?". Each question seems legitimate in its own right, yet the answers that they suggest appear to point into different directions. While the first goal suggests a policy that removes the causes of terrorism, the second goal suggests a policy that fights terrorists. Can a policy meet both goals at the same time? And which of two policies is better if each one is superior to the other with regard to one of the goals?

A model about terrorism prevention should ideally address both goals. If it addresses only one goal, it may favour a policy that achieves this goal at the expense of

[1] This research is supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government. Earlier versions of the paper were presented at the International Summerschool "Philosophy, Probability, and the Special Sciences" (Konstanz, July 2003), the 2nd Conference of the European Consortium for Political Research ECPR (Marburg, September 2003), the workshop "Philosophy and Probability" at the GAP 5 (Bielefeld, September 2003), and the conference "Coercive Power and its Allocation in the Emerging Europe" of the European Center for the Study of Public Choice (ECSPC) and the Interdepartmental Center for International Economics (CIDEI) (Rome, September 2003).
Address for correspondence: Center For Junior Research Fellows, University of Konstanz, 78457 Konstanz (Germany). Telephone: ++49 (0)7531 88-4733. Email: Franz.Dietrich@uni-konstanz.de. Web: www.uni-konstanz.de/ppm/Dietrich.

the other goal: The policy may achieve a high "capture success" while creating even more new terrorists, or create few terrorists while leaving society entirely exposed to them. The purpose of this paper is to present a general method for modelling the combination of both goals. This method consists in

(a) selecting a single objective function that simultaneously reflects both goals;

(b) decomposing this objective function into "inclination parameter(s)" that represent (the degree of) inclination towards terrorism (first goal) and "neutralisation parameter(s)" that represent (the degree of) neutralisation of terrorists' power (second goal) (and possibly other parameters such as policy cost parameters; see Section 1.1);

(c) separately estimating policy effects on each of these parameters;

(d) deducing the overall policy effect on the objective function.

While steps (a) and (b) are a question of model choice, step (c) is surely the most demanding one. Although it is desirable that such an estimation be backed up by empirical data, the limits of a purely statistical estimation will be pointed out. To illustrate the method (a)-(d), I shall apply it to a particularly simple objective function in (a) and decomposition in (b). The objective function to be minimised is simply the *number of free terrorists*, and the decomposition explains this quantity in terms of a single inclination parameter, the overall *number of terrorists*, and a single neutralisation parameter, the *detention rate* or proportion of detained terrorists.[2] Surely, this macro-model is very simple, but it seems well suited for illustrating the general method. (A more elaborate decomposition will also be discussed.)

The difference to game-theoretic models. Unlike most other technical models, this model (that is, the simple example illustrating the method (a)-(d)) is not game-theoretic. While game-theoretic models ask what (rational) actions are made by given terrorists (and governments), my model will ask (among other things) whether and when a person becomes a terrorist in the first place – so the question is not how policies affect a person's (strategic) actions given "terroristic" preferences, but how policies affect the preferences in the first place.[3] While game-theoretic models ignore the question of the formation of "terroristic" preferences or motives and focus on minimising the damage created by given terrorists with these preferences, my model will ignore the different (and differently damaging) actions of terrorists and focus on reducing the number of terrorists in action. So, the difference is that, of the two goals mentioned earlier, my model also addresses the first one but is less elaborate on the second one. This difference might be the source of why game-theoretic models often suggest rather tough policies, while my model perhaps suggests a rather cautious policy response. It would be an interesting challenge to combine both approaches by modelling policy effects on both preferences and actions, i.e. both whether people develop terroristic motives and if so what action they take. This could for instance be

[2] The word "terrorism" will be used as a purely technical term to denote a non-governmental use of force or violence against human or non-human targets. Related words ("terrorist", "terroristic", ...) are used accordingly. Since these are definitions, and not claims, I do not intend to take a stand in the controversy of what terrorism is and who is a terrorist in a non-technical sense.

[3] At first sight, one is tempted to compare "(not) being a terrorist" in my model with "(not) performing terroristic actions" in a game-theoretic model. However, a substantial difference is that the non-terrorist does not like, say, the plane to crash, while the terrorist prefers the plane to crash but may for strategic reasons opt to leave the plane in safety. Formally, terrorists and non-terrorists have different preferences, i.e. assign different utilities to the same outcomes.

achieved by using the method (a)-(d) with neutralisation parameters some of which reflect the power of terrorists in action.

Structure of the paper. I shall illustrate (in Sections 1 and 2) and assess (in Section 3) the general method (a)-(d) by means of the simple example indicated above. More precisely, in Section 1, I present a general taxonomy of policies, and policy effects on the level of inclination (towards terrorism) and neutralisation (of terrorists' power). In Section 2, I focus on symptom-related policies, and discuss the typical effects of their *toughness level* on the level of inclination and neutralisation. For instance, I argue that inclination is often a *U-shaped* function of the toughness level of the capture policy: Inclination is high both for very low and very high toughness, due to too little deterrence resp. too strong a hate effect. In summary, the analysis will suggest that, typically, the *capture policy* should be not very tough and *length of detention* not very short. In Section 3, a methodological assessment of my approach is presented: Is it justified to decompose the objective function and estimate separately policy effects on each inclination and neutralisation parameter, rather than estimating at once policy effects on the objective function? The answer, I argue, depends on the estimation method. Often, a *purely statistical* estimation is impossible, so that the estimation has to be *theory-based*, founded on (ideally empirically/statistically tested) arguments on the causal mechanisms creating policy effects. This approach does indeed benefit from a decomposition, which cuts a complex estimation problem into more manageable ones.

The relation to the literature. To position the model (i.e., the simple illustration given of method (a)-(d)) relative to the literature, let me first note some general characteristics of it. The model is a *macro-model*: It considers the a population as a whole rather than modelling the behaviour of single persons (although the model may to some extent be backed up on the micro-level). Further, the analysis I give of the model is theoretical rather than statistical, yet not game-theoretic, resulting in qualitative, not quantitative, claims and policy recommendations.[4] By contrast, the formal literature on terrorism often takes a game-theoretic or statistical approach; and it typically focusses on a recent event or a particular policy choice problem, whereas my model is general. However, none of these characteristics of my model are necessary features of models following the method (a)-(d).

This paper may be considered either, in view of the specific topic, in the context of the terrorism literature, or, from a methodological point of view, in the context of mathematical and probabilistic modelling in political science. Regarding the specific terrorism literature, my model is little related to previous models or methods by the above remarks. Some of the previous studies take a mainly political and sociological approach; e.g. Dumas (2002), Wulf et al. (2003). Other studies derive policy recommendations based on economic arguments (e.g. Congleton (2002), Frey and Luechinger (2003)), or on game-theoretic modelling (e.g. Brophy-Baermann and Conybeare (1994), Pape (2003), Sandler et al. (1983), Sandler and Enders (2002)). From a methodological perspective, the parameters of my model may be interpreted as macropolitical variables describing the state of society with regard to terrorism. This relates to an approach to political science borrowed from economics, promoted by Cioffi-Revilla (e.g. Cioffi-Revilla (1998, 1998a, 1985), Cioffi-Revilla and Starr (1995)).

[4] I.e., I shall discuss the *direction* in which – but not the exact *extent* to which – policies affect the parameters of interest.

Here, the state of society is represented by a set of macropolitical variables (e.g. immigration rate, social cohesion measures, duration of a coalition, time of a terrorism attack, number of terrorists, etc.), the interactions between which are analysed. Variables may be exogenous or endogenous, more or less easily measurable (observable), and more or less easily predictable. The macrovariables of the present analysis are of the mostly non-measurable type, since terrorists try to appear as non-terrorists. More generally, there is a growing literature in political science that, like this paper, has a probabilistic motivation. Recent examples are probabilistic approaches to the discursive dilemma (e.g. Bovens and Rabinowicz (2001), List (2003), Pettit (2001)); more traditional examples are the probability of cyclic collective preference (e.g. Jones et al. (1995), Tangian (2000)), and approaches to epistemic democracy from the Condorcet jury theorem (e.g. Grofman and Feld (1988), List and Goodin (2001)).

1 A model of cause-related and symptom-related policies

I begin, in Section 1.1, by introducing more formally the simple model used to illustrate the general technique (a)-(d). I then mention the different types of policies and policy effects on the parameters of interest (Section 1.2), followed by some comments about the endogeneity of the number of terrorists (Section1.3) and a heuristic comparison between the potential of cause- and symptom-related policies (Section 1.4). Optimal polices in my model are formally defined in Section 1.5. In Section 1.6, alternative decompositions of F are discussed, to illustrate the broader scope of the method (a)-(d).

1.1 A simple objective function and its decomposition

I assume that a *terrorism prevention policy* or just *anti-terrorism policy* has to be devised by some governmental or supernational body or organisation. The terrorism threat arises from (known or unknown) members of a given group of people. This group may be defined geographically, ethnically, economically or in any other way.[5] The group may consist of part or all of the domestic population (national anti-terrorism policy), or contain foreign population (international anti-terrorism policy), perhaps even the world's entire population (beside the policy-makers and -implementers). Throughout, the words "person", "individual" and "people" refer to persons from that group. To simplify, at any given time each person is in any one of the following three possible states *"non-terrorist"*, "*detained terrorist*", and "*free terrorist*", without further refinement. Note that, by assumption, non-terrorists are never detained. Policies have causal effects on people's states, which I shall explore. For instance, a tough capture policy increases the probability that terrorists are detained; and it may or may not increase the overall number of terrorists depending on whether deterrence or hate effect dominates. Each policy leads to a certain *group composition* (or *group structure*) given by the proportion taken by each of the three subgroups. This group composition can be given two interpretations: For the *finite*

[5] Of course, the group may change over time, due to births, deaths, migrations, etc.

time horizon interpretation, it is the group composition as given after the policy has been in action for a given fixed period of time (e.g. after 5 years); for the *(long-term) equilibrium* interpretation, it is the (equilibrium) group composition to which, ceteris paribus, the policy has a tendency to lead to in the long run.

The objective function. By assumption, the only policy goal is to[6]

minimise the resulting number F of free terrorists,

regardless of how the rest of the group is split up into non-terrorists and detained terrorists. Note that the strategy of reducing F to 0 by simply capturing *everybody* is excluded since non-terrorists are, by assumption, never detained.[7] Further, note that this objective function neglects two important aspects:

1. It neglects *costs* of anti-terrorism policies, including "costs" of human lives (on all sides), financial costs, costs of quality of life (due to restrictions of individual liberties, military presence, etc.). In what direction does this distort the picture? In the case of *capture policy*, since all of these costs plausibly seem to be increasing functions of the toughness level, their additional inclusion into the objective function would lead to a less tough optimal capture policy. In this sense, my objective function is biased towards higher toughness of the capture policy. So, insofar as my analysis will suggest a little tough capture policy, this conclusion would get further reinforced by including costs.[8]

2. It neglects that the terrorism threat need not be proportional to the number of free terrorists. This neglects that some terrorists may be more dangerous than others, that a critical number of terrorists might be necessary to execute certain large attacks, and that terrorist networks play an important role; and, last but not least, it neglects what game-theoretic models focus on, namely that a free terrorist may be more threatening given certain policies than given others.

Decomposition of F. The perhaps simplest decomposition of F explains F by two factors: overall number of terrorists and detention rate. To explain, let the group's composition (in response to the policy) be given by the number of free terrorists F, the number of detained terrorists D, and the number of non-terrorists N; the overall number of terrorists is then the sum $T := F + D$, and the detention rate is the ratio $R = D/T = D/(F + D)$. The minimand F can be expressed in terms T and R:

$$F = (1 - R)T,$$

[6] This is the same as minimising the proportion resp. probability of free terrorists, as done in the abstract. For ease of exposition, however, I prefer to talk in terms of *absolute numbers* rather than *proportions* resp. *probabilities*.

[7] In principle there are two ways to exclude undesired solutions such as "capturing everybody": Either the minimand is defined so as to become large on the undesired policy "capture everybody" – which is not the case for F but would be the case if costs were included – or, as I prefer to do here, the decision space of allowed policies is a priori restricted so as not to contain certain policies (such as "capturing everybody") and the minimisation is then performed within this restricted domain.

[8] Including costs means minimising not F but the sum $F' := F + C$ where C denotes costs of all sorts (measured in a suitable unit to justify adding C to F). To apply the method (a)-(d), one has to decompose the new objective function F'. A decomposition of F' may be obtained as the sum of a decomposition of F (e.g. $(1 - R)F$) and a decomposition of C (i.e. a sum of the different types of costs). One then has to estimate policy effects on each inclination, neutralisation, and cost parameter of the decomposition of F'. My analysis of F can be viewed as a contribution to the study of the first term, F, of the ("true") objective function F'.

because $F = (F/T)T$ where $F/T = (T - D)/T = 1 - R$. This decomposition shows exactly how, by influencing the detention rate R and the number of terrorists T, the policy may influence the number of free terrorists F. Note that the relation $F = (1 - R)T$ is perfectly equivalent to the probabilistic relation P("free terrorist") = P("free" | "terrorist")P("terrorist") mentioned in the the abstract.[9]

1.2 The different parts of an anti-terrorism policy

Given the aim of minimising the number of free terrorists $F = (1 - R)T$, let us call "anti-terrorism policy" the collection of *all measures that affect one or both of T and R* (and thereby affect F). Since $F = (1 - R)T$ is increasing as a function of T and decreasing as a function of R, it is desirable in principle (but hard to achieve) that the policy simultaneously makes number of terrorists T small and detention rate R large. I subdivide an anti-terrorism policy into the *cause-related policy* which affects only T and the *symptom-related policy* which affects both R and T (Figure 1).[10]

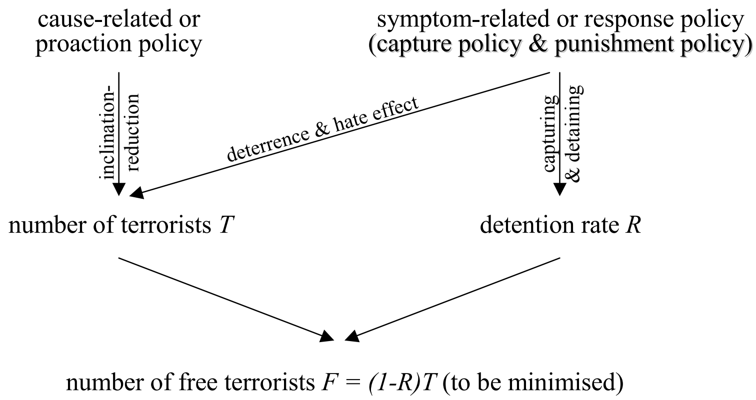

Figure 1: Causal effects of both parts of an anti-terrorism policy on T and R, and hence on F.

The cause-related policy. This part of an anti-terrorism policy consists of all measures aiming to reduce people's inclination towards terrorism (except through deterrence, which is part of the symptom-related policy described below). Such measures are proactive rather than responsive: They try to prevent people from being terrorists in the first place, without helping to capture existing terrorists. So they reduce the number of terrorists T without affecting the detention rate R (by the "idealising assumptions" below). Such measures may consist in raising the standard of living, fighting poverty, improving the level of education, etc. If the group includes population of foreign countries, the cause-related policy may include the kind and extent of political and economical relations maintained with these countries, the kind

[9] After dividing both sides of $F = (1 - R)T$ by the group size m $(= N + F + D)$, the relation becomes $F/m = (1 - R)(T/m)$, where F/m is the proportion of free terrorists or probability P("terrorist"&"free"), T/m is the proportion of terrorists or probability P("terrorist"), and $1 - R$ is $1 - P$("detained"|"terrorist") = P("free"|"terrorist").

[10] The names "cause-related" and "symptom-related" are not meant to express an a priori bias in favour of cause-related policies.

and extent of development policy in the case of third world countries, and so on.

The symptom-related policy. This part of an anti-terrorism policy is responsive rather than proactive: It consists of all measures aimed at reducing or "neutralising" the threat of terrorism *once there are terrorists.* A reasonable subdivision seems to be into the "*capture policy*", i.e. the kind and extent of efforts to capture terrorists, and the "*punishment policy*", i.e. the kind and extent of punishment of captured terrorists, in turn subdividable into *length of detention* and *conditions of detention.* (Given the sole aim of minimising F, I neglect the "*defence policy*" against attacks, examples of which are defence installations on the roof of a skyscraper and the US plans of a "missile defence shield".) Crucially, the symptom-related policy affects both detention rate R and number of terrorists T (Figure 1). The typical effects on R and T are analysed in Section 2. To anticipate, more toughness typically increases detention rate R; and whether it reduces or increases people's inclination towards terrorism and hence the number of terrorists T depends on whether deterrence or the hate effect dominates. If the group is or belongs to the population of the domestic country, the symptom-related policy is determined by factors such as the kind and extent of police presence and control (capture policy) and the judicial system (punishment policy). In an "international anti-terrorism policy", the capture policy might include secret services, military support given to certain governments, or military interventions; a current example of punishment policy is the intended creation of an international war crime tribunal.

Idealising assumptions. This and the following analysis is based on an idealisation or simplification of the real world in which "everything affects everything", and often in a number of complex ways. This idealisation is that, by assumption, number of terrorists T reflects only the group's (degree of) *inclination* towards terrorists, and that detention rate R reflects only the (degree of) *neutralisation* of the terrorism threat. Of course, this is not to say that T and R reflect *all* information about inclination resp. neutralisation – for instance T does not tell how much inclined the different age groups are –, and so T and R are only certain ways of *summarising* inclination resp. neutralisation into a single quantity. The idealisation is that T summarises information *only* about inclination, and R information *only* about neutralisation. Such assumptions are not implausible, but a close look at them reveals that they are far from cogent.[11] So, a non-idealised analysis would have to consider effects of the cause-related policy on detention rate R, and effects of the symptom-related policy on T in addition to deterrence and hate effect. But a better strategy, if the idealisation is considered as problematic, is probably to choose a different decomposition of F (see Section 1.6).

1.3 Endogenous number of terrorists

As mentioned earlier, models of terrorism prevention differ, among other things, regarding whether they treat the number of terrorists as exogenously given (as in most

[11] Here is a counterexample for R. Imagine that a certain policy leads to a higher inclination (towards terrorism) among young people, and let me show how this raises the detention rate R. Given that the average age at which a person becomes a terrorist is lower, the average terrorist passes a longer proportion of his or her "life as a terrorist" in prison (assuming that captured terrorists are detained for the rest of their lives – a longer period for younger captured terrorists). This raises the detention rate R, and so R partly reflects how inclination is distributed across age groups.

game-theoretic models) or as endogenous and influenced by the policy (as here). If the number of terrorists T were in fact exogenous, there would be neither the possibility of a cause-related anti-terrorism policy, nor a reason to worry about "side effects" of a symptom-related policy on the number of terrorists, of the form of deterrence and hate effects. In the product $F = (1 - R)T$ the term T would be a constant, so that minimising F would reduce to minimising the first factor $1 - R$, i.e. maximising the detention rate R – which would clearly suggest a tough capture policy. But this paper treats T as endogenous and hence rejects $1 - R$ (resp. R) as an objective function. The reduction of each factor $1 - R$ and T is important in its own right and deserves full attention as a policy goal.

1.4 The relative potential to reduce the number of terrorists T and increase the detention rate R

One might already now wonder where the potential of reducing the number of free terrorists F is higher: Is it through increasing the detention rate R (by a tougher symptom-related policy) or through lowering the number of terrorists T (by an improved cause-related policy and a "well-tuned" symptom-related policy, as argued later)? No easy answer is possible, but a heuristic argument points towards a high potential of reducing T. The argument is that the proportion of terrorists varies dramatically between societies: It may easily be ten, hundred or more times higher in societies with a high level of instability, poverty and "brutalisation", such as Somalia or the Gaza Strip, than in relatively stable and peaceful societies such as France and the United States of America. By contrast, such variations are unrealistic for the factor $1 - R$. Given that T seems more variable than $1 - R$, the potential of reducing number of free terrorists F through reducing T might exceed that through raising R. For example, a raise of R from, say, 0.4 to, say, 0.7 (although possibly involving a permanent worldwide pursuit of terrorists) implies, for fixed T, no more than a drop of $F = (1 - R)T$ from $(1 - .4)T = 0.6 \times T$ to $(1 - .7) = 0.3 \times T$, i.e. a drop by 50%.

1.5 Optimal policies

For later purposes I need to introduce some notation and formalise the policy choice problem as defined in Section 1.2. An anti-terrorism policy is a pair (χ, σ), where χ denotes a cause-related policy and σ denotes a symptom-related policy. Let $\boldsymbol{\chi}$ be the set of all cause-related policies χ considered, and $\boldsymbol{\sigma}$ the set of all symptom-related policies σ considered. Recalling that the detention rate R is affected only by the symptom-related policy σ, while the number of terrorists T is affected by the entire anti-terrorism policy (χ, σ) (Figure 1), I denote by

- $R(\sigma)$ the detention rate for symptom-related policy σ,
- $T(\chi, \sigma)$ the number of terrorists T for anti-terrorism policy (χ, σ),
- $F(\chi, \sigma) = (1 - R(\sigma))T(\chi, \sigma)$ the number of free terrorists F for anti-terrorism policy (χ, σ).

By definition, an anti-terrorism policy (χ^*, σ^*) is optimal if it minimises $F(\chi, \sigma)$, i.e. if[12]

$$F(\chi^*,\sigma^*) = \min_{(\chi,\sigma)} F(\chi,\sigma), \text{ resp. } (1-R(\sigma^*))T(\chi^*,\sigma^*) = \min_{(\chi,\sigma)}\{(1-R(\sigma))T(\chi,\sigma)\},$$

where in these minima (χ,σ) ranges over the set of all considered anti-terrorism policies, i.e. over the set $\boldsymbol{\chi}\times\boldsymbol{\sigma}$ (of all pairs (χ,σ) where $\chi\in\boldsymbol{\chi}$ and $\sigma\in\boldsymbol{\sigma}$) if the choice of χ and σ do not constrain each other, and some subset of $\boldsymbol{\chi}\times\boldsymbol{\sigma}$ in the presence of (budget) constraints.[13]

Narrower choice problems. In practice, however, one is rarely confronted at once with the simultaneous choice of the entire anti-terrorism policy (χ,σ), partly because choice problems come "bit by bit" and responsibilities for different policy parts lie in different hands[14]. First, consider the (still very large) problems of choosing just the symptom-related policy $\sigma\in\boldsymbol{\sigma}$, or just the cause-related policy $\chi\in\boldsymbol{\chi}$. When just σ is chosen, χ is fixed, and hence F and T are functions of σ only; by definition, symptom-related policy $\sigma^*\in\boldsymbol{\sigma}$ is optimal if it minimises $F(\sigma)$, i.e. if[12]

$$F(\sigma^*) = \min_{\sigma\in\boldsymbol{\sigma}} F(\sigma), \text{ resp. } (1-R(\sigma^*))T(\sigma^*) = \min_{\sigma\in\boldsymbol{\sigma}}\{(1-R(\sigma))T(\sigma)\}. \quad (1)$$

Correspondingly, when choosing just χ, then σ is fixed, and so I now drop the dependence of F,T,R on σ; by definition, cause-related policy $\chi^*\in\boldsymbol{\chi}$ is optimal if it minimises $F(\chi)$, i.e. if[12,15]

$$F(\chi^*) = \min_{\chi\in\boldsymbol{\chi}} F(\chi), \text{ resp. } T(\chi^*) = \min_{\chi\in\boldsymbol{\chi}} T(\chi). \quad (2)$$

Now assume the choice problem refers to only *some part of* the symptom- resp. cause-related policy, for instance the length of detention of captured terrorists, or the size of prison cells, or the size of troops in an occupied country, or the amount of development aid for one particular country, all of which are only subcomponents of σ resp. χ. For ease of notation, I still use the symbols σ and χ for policy options (and the symbols $\boldsymbol{\sigma}$ and $\boldsymbol{\chi}$ for decision spaces) even when the decision concerns only a subcomponent of the symptom- resp. cause-related policy. For instance, a $\sigma\in\boldsymbol{\sigma}$ may stand just for a possible length of detention, and a $\chi\in\boldsymbol{\chi}$ just for a possible amount of development aid for one particular country. With this notation, an optimal policy is still given by the condition (1) resp. (2).

1.6 Excursion: some alternative decompositions of F

The method (a)-(d) is not restricted to the decomposition $F=(1-R)T$. In fact, a disadvantage of this decomposition is that policy effects on T and R may be complex and not fully transparent, hence hard to estimate (if the estimation is theory-based).

[12] More comprehensive optimality criteria would include costs, as mentioned in the introduction.

[13] There may, for instance, be the budget constraint of a certain maximal sum-total amount of money spent on the policy pair. Then certain policy pairs $(\chi,\sigma)\in\boldsymbol{\chi}\times\boldsymbol{\sigma}$ are excluded, since for instance the choice of an "expensive" χ requires that of a "little expensive" σ.

[14] Capture policy might be influenced significantly by the ministry of defence, punishment policy by the ministry of justice, and cause-related policy by the foreign ministry and the ministry for development policy.

[15] Here the factor $1-R$ is not needed on both sides of $T(\chi^*) = \min_{\chi\in\boldsymbol{\chi}} T(\sigma)$ since $1-R$ is independent of χ.

One may consider other decompositions of F as superior in which F is expressed (explained) in terms of inclination and neutralisation parameters that are easier to analyse than T and R. For instance, in one of the dynamic models in Dietrich (2004), the *equilibrium* number of free terrorists is found to be given by[16]

$$F = \frac{m a_{\mathrm{nf}}}{a_{\mathrm{fn}} + a_{\mathrm{fd}} + a_{\mathrm{nf}} + b + a_{\mathrm{nf}} a_{\mathrm{fd}}/b}, \tag{3}$$

where m is the group size, b is the birth rate, and $a_{\mathrm{nf}}, a_{\mathrm{fn}}, a_{\mathrm{fd}}$ are, respectively, the frequency of non-terrorists becoming free terrorists, of free terrorists (re)becoming non-terrorists, and of free terrorists becoming detained terrorists, i.e. being captured. The parameters $a_{\mathrm{nf}}, a_{\mathrm{fn}}, a_{\mathrm{fd}}$ describe the amount of "flow" between the three categories of people, and, more precisely, the percentage of persons moving from one category to another per time step (year, month, etc.). a_{fd} plays the role of "neutralisation parameter" (like R) since it reflects how many free terrorists are captured in a time step; so that a_{fd} is affected by the symptom-related policy σ only: $a_{\mathrm{fd}} = a_{\mathrm{fd}}(\sigma)$. a_{nf} and a_{fn} play the role of "inclination parameters" (like T) since they reflect how many people decide in a time step to become terrorists resp. to re-become non-terrorists; so that a_{nf} and a_{fn} are affected by the cause-related policy χ, and by the symptom-related policy σ through deterrence and hate effect: $a_{\mathrm{nf}} = a_{\mathrm{nf}}(\chi, \sigma)$ and $a_{\mathrm{fn}} = a_{\mathrm{fn}}(\chi, \sigma)$. Note that, not surprisingly, F is decreasing as a function of a_{fd} (reflecting that more captures lower F) and increasing resp. decreasing as a function of a_{nf} resp. a_{fn} (reflecting that higher inclination towards terrorism raises F). The potential advantage of the decomposition (3) over the decomposition $F = (1-R)T$ is that policy effects on $a_{\mathrm{nf}}, a_{\mathrm{fn}}, a_{\mathrm{fd}}$ may be easier to analyses than those on R, T. For it may be more transparent how particular levels of inclination/neutralisation (created by policies) translate into particular values of the parameters. But the decomposition $F = (1-R)T$ is much simpler, and hence chosen here for illustrating the method (a)-(d).

Instead of the *equilibrium* decomposition 3, one may also choose a *finite-time-horizon* decomposition expressing the number of free terrorists F after, say, five years, in terms of the parameters $a_{\mathrm{nf}}, a_{\mathrm{fn}}, a_{\mathrm{fd}}$, m and b, and, in addition, the initial numbers of free and detained terrorists. Further, the model may be made more complex by using an (equilibrium or finite-time-horizon) decomposition of F in terms of more or other inclination and neutralisation parameters than $a_{\mathrm{nf}}, a_{\mathrm{fn}}, a_{\mathrm{fd}}$, for instance additional parameters a_{dn} and a_{df} by allowing that detained terrorists are released from prison.

2 The symptom-related policy and its toughness level

In this section I exclusively consider the choice of (some part of) the symptom-related policy, and more precisely of the *toughness level* of the latter. I begin by introducing the term "toughness level" (Section 2.1), which allows to consider the parameters F, T, R as functions of the toughness level (Section 2.2). Then, I discuss the typical shape of these functions (Sections 2.3 and 2.4), with resulting policy recommendations summarised in Section 2.5. In Section 2.6, two stereotypical cases are discussed, in

[16] The formula contains no "a_{dn}" and "a_{df}" since it assumes among others that captured terrorists are never released.

which minimising $F = (1 - R)T$ essentially reduces, in the first case, to minimising the first factor $1 - R$, achieved by maximal toughness, and, in the second case, to minimising the second factor T, achieved by creating a climate in which inclination towards terrorism is minimal.

2.1 Toughness level

For the sake of numerical representation, let us assign to each policy $\sigma \in \boldsymbol{\sigma}$ a numerical value $\tau = \tau_\sigma \in \mathbf{R}$ representing the "toughness level" of that policy, where a tougher policy is, of course, assigned a higher toughness τ_σ. Some examples:

- If the choice is that of the *length of detention*, toughness τ of the policy might be defined as average number of years a captured terrorist is detained.

- If the choice relates to *conditions of detention*, toughness τ might, depending on the specific choice problem, be defined as the number of hours of work per day per detainee, or as minus size of prison cells in square meters (a "minus" because smaller cells should be assigned higher toughness τ), etc.

- If the choice is that of the amount of airport controls, toughness τ might be defined as number of airport controllers, or number of security checks per passenger, or overall time the security check takes per passenger, etc.

- If the choice is that of the amount of military presence in an occupied country with the mission of locating terrorists, toughness τ might be defined as number of soldiers stationed, or as financial costs of the military presence, etc.

As explained below, to justify using a unidimensional toughness measure τ I have to assume a sufficiently narrow policy choice problem, for instance one of the mentioned ones, but probably not the problem of choosing the entire symptom-related policy. (Otherwise, I would have to use a multi-dimensional toughness measure $(\tau_1, ..., \tau_k)$, which I avoid for the sake of simplicity.[17]) Further, as the examples show, in a given problem there may be different (ordinally equivalent[18]) ways of defining toughness. Hence the choice of the toughness measure may involve some amount of arbitrariness.

If the set of policies $\boldsymbol{\sigma}$ is finite – a case I will exclude – the toughness levels τ_σ, $\sigma \in \boldsymbol{\sigma}$, of these policies, are finitely many numbers, somehow scattered over the axis of real numbers $\mathbf{R}$; for instance, $\boldsymbol{\sigma}$ may consist of only three policies, with toughness levels .1, .7 and .9. However, the discussion will benefit from considering an entire *interval* of toughness levels $[\tau_{\min}, \tau_{\max}]$ (with $\tau_{\min} < \tau_{\max}$). Here, toughness $\tau = \tau_{\min}$ stands for the *minimally feasible* toughness level along the relevant dimension: "immediate release after capture" when choosing the length of detention, "no captures" when choosing the capture policy, "no airport controls" when choosing the amount of airport controls, "no soldiers" when choosing the number of soldiers to be stationed in

[17] The components of the toughness vector $(\tau_1, ..., \tau_k)$ represent the toughness levels along the different toughness dimensions: Instead of analysing functions $R(\tau)$ and $T(\tau)$, one would have to analyse functions $R(\tau_1, ..., \tau_k)$ and $T(\tau_1, ..., \tau_k)$.

[18] For instance, one would expect that the number of airport controllers is higher (for one policy than for another), *if and only if* the number of security checks per passenger is higher, *if and only if* the overall time the security check takes per passenger is higher, and so on..Formally, two toughness measures τ_σ and $\tilde{\tau}_\sigma$ $(\sigma \in \boldsymbol{\sigma})$ are *ordinally equivalent* if for any pair of policies $\sigma, \sigma' \in \boldsymbol{\sigma}$, $\tau_\sigma > \tau_{\sigma'}$ if and only if $\tilde{\tau}_\sigma > \tilde{\tau}$; i.e. σ is tougher than σ' according to one of the measures if and only if it is so according to the other measure.

a country, etc. And toughness $\tau = \tau_{\max}$ stands for the *maximally feasible* toughness level along the relevant dimension given limited resources: "never released" when choosing length of detention, "maximal affordable size of troops" when choosing the size of the troops occupying a country, etc. Besides the extreme toughness levels $\tau = \tau_{\min}$ and $\tau = \tau_{\max}$, the interval $[\tau_{\min}, \tau_{\max}]$ contains a continuum of intermediate toughness level $\tau_{\min} < \tau < \tau_{\max}$. Note the idealisation here: In practice, a choice from the (infinite) set of toughness levels $[\tau_{\min}, \tau_{\max}]$ is not only unfeasible since the choice is limited to relatively few policies σ resp. toughness levels τ, but perhaps also unreasonable since toughness levels near the extremes $\tau_{\min}$ or $\tau_{\max}$ are unlikely to be efficient.

When is a unidimensional toughness measure justified? While an exact *cardinal* measure of toughness may be ambiguous – can one say that "policy σ has toughness level 1.5"? – toughness in the sense of an *ordinal* measure seems less problematic since often one *can* say that "policy σ is tougher than policy σ'". So it is important that we interpret the chosen toughness measure merely as an *ordinal* toughness measure, i.e. as one of many possible numerical representations of an underlying toughness ordering of the policies. But even ordinal toughness comparisons may become problematic if policy σ is tougher than policy σ' along some dimension but milder along another dimension. For instance, policy σ might consists in a tougher capture policy but a milder punishment policy. This points towards the fact that toughness level is often a multi-dimensional parameter. Assume, for instance, that the choice is of the entire symptom-related policy. Then toughness is at least two-dimensional: toughness of punishment policy and toughness of capture policy, if not higher-dimensional by further subdividing punishment policy and capture policy. Indeed, punishment policy might be subdivided into *length of detention* and *conditions of detention,* the latter being further subdividable into size of cells, temperature of cells, etc. And, as one of many possible subdivisions, toughness of capture policy may be taken to consist of toughness experienced by the terrorists ("toughness to terrorists") and toughness experienced by the non-terrorists ("toughness to the external world"); a war seems tough both to terrorists and the external world, while isolated "surgical" military interventions might be tough mainly to the terrorists and less so to the external world. So, for the above unidimensional toughness τ to be justified, I have to assume that the policy choice problem is narrow enough so that the toughness of the policies $\sigma \in \boldsymbol{\sigma}$ considered actually varies along only one toughness dimension.[19] It is then always clear which of two policies is tougher, and so the different policies $\sigma \in \boldsymbol{\sigma}$ can be unambiguously ordered in terms of toughness: They may be thought of as ordered

[19] I also allow the "quasi-unidimensional" case in which toughness varies along different dimensions but does so never in opposite directions, so that if a policy is tougher than another along some dimension, it may not be milder along another dimension. The reason for allowing such cases is that one may treat them as unidimensional by collecting all toughness dimensions involved (e.g. length of detention, size of cells, temperature of cells, etc.) into a single broader toughness dimension (e.g. "toughness of punishment policy"). For instance, the choice between "war" and "no war" (against a given country) affects both *toughness to terrorists* and *toughness to the external world*, but it does so in the same direction in both cases since a war is the tougher option both to terrorists and to the external world. By contrast, toughness varies in opposite directions along different dimensions when choosing how to split a fixed budget into one amount spent on more airport controls and one amount spent on more troops in a foreign country: More toughness (money) along one dimension implies less toughness (money) along the other dimension.

on a line ranging from "least tough" to "toughest".[20] My toughness measure is then just one (of many possible) numerical representations of this toughness ordering.

2.2 The parameters F, T, R as functions of toughness

The aim of the following analysis is to examine detention rate R, number of terrorists T, and number of free terrorists F as functions of toughness τ. So, for any toughness $\tau \in [\tau_{\min}, \tau_{\max}]$ let $F(\tau)$, $T(\tau)$ and $R(\tau)$ denote the value taken by, respectively, F, T and R when (the policy of) toughness τ is implemented. The policy optimisation problem (1) may now be expressed in terms of toughness: Toughness $\tau^* \in [\tau_{\min}, \tau_{\max}]$ is optimal if it minimises $F(\tau)$, i.e. if

$$F(\tau^*) = \min_{\tau \in [\tau_{\min}, \tau_{\max}]} F(\tau), \text{ resp. } (1 - R(\tau^*))T(\tau^*) = \min_{\tau \in [\tau_{\min}, \tau_{\max}]} \{(1 - R(\tau))T(\tau)\}.$$

When analysing the typical shapes of the functions $R(\tau), T(\tau), F(\tau)$, it is important to keep in mind that the exact shapes strongly vary from case to case, as they depend on the following factors:

(i) the type of policy choice problem considered (length of detention of prisoners, size of the army occupying a foreign country, etc.);

(ii) the specific group to which the policy is applied (the population of the Middle East, the domestic population, a minority living in the domestic country, etc.), and all exogenous factors determining the behaviour of this group (level of education, cultural heritage, economic situation, sociological factors, climate, etc.);

(iii) the interpretation of the parameters F, T, R: fixed time horizon or long term equilibrium interpretation (see the introduction);

(iv) the specific way toughness is measured, i.e. the way τ is defined.[21]

For the sake of generality, I do not restrict the discussion to particular answers in each of (i)-(iv); consequently, the following analysis of the shapes of the functions $F(\tau), T(\tau), R(\tau)$ is *qualitative*, i.e. concerned with signs of (first and second) derivatives, minima, etc. I shall identify general features of these functions valid in *most* cases, and point out potential exceptions.

Smooth functions. Strictly speaking, the functions $F(\tau), T(\tau), R(\tau)$ would have to be "step functions" rather than being continuous.[22] But they are more easily imagined as continuous functions, and so the discussion and all figures will assume continuity – a matter of "smoothing out" the steps, which is no distortion in a large enough group. When I refer to derivatives, I moreover assume differentiability.

[20] Formally, the "tougher than" relation is a *linear order*, i.e. a complete, transitive and irreflexive order.

[21] Recall that there are many ordinally equivalent toughness measures (which are strictly increasing transformations of each other). If τ is replaced by, for instance $\bar{\tau} := \tau^3$, then R may be a quite different function of $\bar{\tau}$ than of τ. More precisely, the signs of the first derivatives of $R(\tau)$, $T(\tau)$ and $F(\tau)$ are invariant under transforming the toughness measure (by the chain rule for differentiation the composition of two functions), but the second derivative is not invariant. So, the claims about first derivatives (viz. that, typically, $R(\tau)$ is increasing and $T(\tau)$ is U-shaped) do not depend on the chosen toughness measure, whereas statements about second derivatives (viz. the typical concavity of $R(\tau)$ and convexity of $T(\tau)$) do not enjoy independence of the chosen toughness measure.

[22] The functions $F(\tau)$ and $R(\tau)$ represent numbers of persons and hence "jump" between integer numbers, and detention rate $R(\tau)$ represents a fraction of the number of terrorists and hence "jumps" between fractional numbers.

2.3 Effect of toughness on detention rate: function $R(\tau)$

I begin with the easiest function, namely detention rate $R(\tau)$.As illustrated by Figure

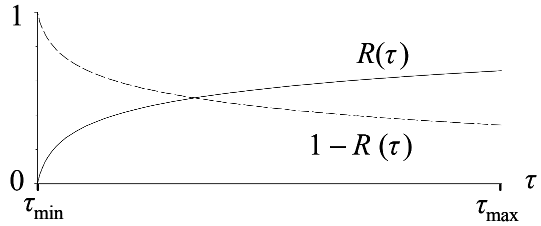

Figure 2: Example of detention rate $R(\tau)$ as a function of toughness τ and corresponding function $1 - R(\tau)$

2, this function is typically (strictly) increasing and (strictly) concave[23]; in other words, $R(\tau)$ has a positive first and a negative second derivative; in even other words, toughness τ has positive and decreasing returns on R. Indeed, more toughness τ plausibly allows to detain a larger proportion of terrorists (higher $R(\tau)$), but a given marginal increase of toughness $\Delta\tau$ should have less effect on $R(\tau)$ if τ resp. $R(\tau)$ is already high since raising R is the harder, the higher R already is. Let me illustrate this by taking up some of the earlier examples. First, the claim that $R(\tau)$ is an increasing function is confirmed by the following examples: If toughness τ is average number of years for which captured terrorists are detained, then higher τ entails higher detention rate $R(\tau)$; if toughness τ is number of security checks at the airport, or size of an army in an occupied country, then higher τ entails more "capture success", hence a higher detention rate $R(\tau)$. In these examples the claim that $R(\tau)$ is concave (has a negative second derivative) seems also reasonable. But again, neither increase nor concavity of $R(\tau)$ are fully obvious in these (and other) examples, and for this reason one might prefer the decomposition of Section 1.6 in which it seems more clear that the neutralisation parameter a_{fc} is increasing and concave as a function of the toughness of capture policy.

In Figure 2, the function $R(\tau)$ has two other features beside of being increasing and concave: it starts at 0 for minimal toughness $\tau = \tau_{\min}$, but does not reach 1 even at maximal toughness $\tau = \tau_{\max}$; in other words, no terrorist is detained at minimal toughness, but not every terrorist is detained at maximal toughness. The latter feature ($R(\tau_{\max}) < 1$) is generally true: Whatever the policy choice problem, even maximal toughness cannot create a situation where not a single terrorist is free. The former feature ($R(\tau_{\min}) = 0$), however, is not general and may or may not hold depending on the type of policy choice problem considered. First, two examples where $R(\tau_{\min}) = 0$ does indeed hold: If toughness τ is length of detention then minimal toughness $\tau_{\min}$ means detention length of zero, so that there are no detained terrorists; and if the choice problem is the (very broad) one of choosing the capture policy, then minimal toughness $\tau_{\min}$ means "no efforts to capture", so that again

[23] A function $g(\tau)$ ($\tau \in [\tau_{\min}, \tau_{\max}]$) is *concave* (*convex*) if, for all $\tau, \tau' \in [\tau_{\min}, \tau_{\max}]$ and all $\lambda \in (0, 1)$, $g(\lambda\tau + (1-\lambda)\tau') \geq (\leq) \lambda g(\tau) + (1-\lambda) g(\tau)$. The definition of *strictly* concave (convex) is obtained by replacing the weak inequality $\geq$ ($\leq$) by the strict inequality $>$ ($<$).

there are no detained terrorists. But these two examples may in fact be the only ones in which $R(\tau_{\min}) = 0$. Indeed, I have $R(\tau_{\min}) > 0$ when the choice relates to a *condition* (rather than the *length*) of detention (as discussed below), or to only a *part* of capture policy (since minimal toughness for this part does not prevent captures coming from other parts of capture policy[24]).

Exceptions. There is an important class of policy choice problems for which the function $R(\tau)$ is typically *flat*, i.e. where toughness has no effect on detention rate. This class consists of all choices of *conditions of detention*, i.e. of all subdimensions of punishment policy other than length of detention. Indeed, the proportion of detained terrorists $R(\tau)$ does not increase by adding toughness of any kind to "life in prison" (more hours of work, smaller cells, etc.) without increasing the length of detention.

2.4 Deterrence effect and hate effect of toughness on number of terrorists: function $T(\tau)$

I now turn to the typical shape of the function $T(\tau)$, i.e. to the causal effect of toughness τ on number of terrorists T. I shall argue that this effect is the result of deterrence and the so-called hate effect, the combination of which typically leads to a *U-shaped* function $T(\tau)$ as in Figure 3 – again with important exceptions in the field of punishment policy choice. I call a function (of $\tau \in [\tau_{\min}, \tau_{\max}]$) "U-shaped" if it is first strictly decreasing and then increasing, i.e. strictly decreasing on $[\tau_{\min}, \tau_0]$ and strictly increasing on $[\tau_0, \tau_{\max}]$ for some intermediate toughness $\tau_0 \in (\tau_{\min}, \tau_{\max})$. The function $T(\tau)$ might in addition be convex[23] (as in Figure 3 and as a "U" suggests), in which case I talk of "convex U-shapedness". I call "deterrence" the reduction of T due to fear of capture and punishment. "Hate effect" is used (by a stretch of language) as a collective term to denote the increase of T due to all sorts of (terrorism-proneness-increasing) feelings or attitudes brought about by (the toughness of) the policy: hate against those fighting terrorism, pride, solidarity with other terrorists, general "brutalisation" of society, etc.[25]

Optimal toughness. Before proceeding to the justification of the U-shapedness of $T(\tau)$, it is important to recall that I aim to minimise not number terrorists $T(\tau)$ but number of *free* terrorists $F(\tau)$, the lower curve in Figure 3. The function $F(\tau)$ arises by multiplying $T(\tau)$ with the function $1 - R(\tau)$, the dashed curve in Figure 2. By the last subsection, $R(\tau)$ is a concave increasing function, and so $1 - R(\tau)$ is a convex decreasing one. If it is true that $T(\tau)$ is U-shaped, multiplying $T(\tau)$ with the convex decreasing function $1 - R(\tau)$ "shifts" the "U" towards the bottom-right, as shown in Figure 3. If $T(\tau)$ grows sufficiently strongly for high toughness τ (i.e. if the upturn of the "U" is sufficiently pronounced), multiplication with $1 - R(\tau)$ preserves U-shapedness, i.e. the resulting function $F(\tau) = (1 - R(\tau))T(\tau)$ is a second U-shaped function rather than being a "curved L"-shaped; $F(\tau)$ differs from $T(\tau)$ in that it first falls at a higher rate than $T(\tau)$ and then increases at a smaller rate than $T(\tau)$[26]. The optimal toughness τ_{optimal}, which minimises $F(\tau)$, is strictly higher

[24] For example, the minimally tough policy of "no airport controls" excludes that terrorists are captured when they enter a plane, but terrorists may still be captured at other occasions given some toughness along other dimensions (secret service, hidden cameras, police force, etc.).

[25] It is irrelevant here whether it can in any sense be legitimate or rational to develop such feelings in response to the policy; what matters is whether and to what extent these feelings are developed.

[26] In other words, on the entire interval $[\tau_{\min}, \tau_{\max}]$ the *growth rate* (this term is defined in footnote

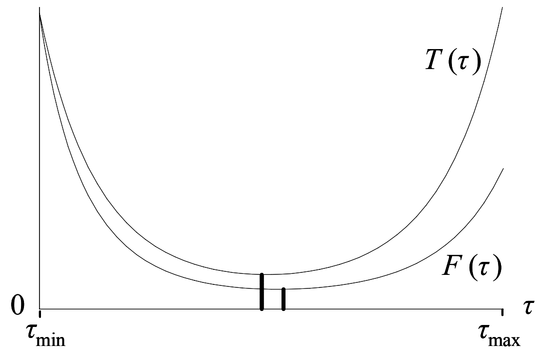

Figure 3: Example of function $T(\tau)$ and resulting function $F(\tau) = (1 - R(\tau))T(\tau)$; minima of $T(\tau)$ and $F(\tau)$ indicated by thick lines

than the toughness τ_0, which minimises $T(\tau)$. In other words, the policy should be tougher than what minimises the number of terrorists $T(\tau)$. The amount by which the policy should be tougher, i.e. by which τ_{optimal} (the minimum of $F(\tau)$) exceeds τ_0 (the minimum of $T(\tau)$), depends on the rate at which $T(\tau)$ grows (and $1 - R(\tau)$ falls) to the right of τ_0: The optimal toughness τ_{optimal} is only little higher than τ_0 if the upturn of $T(\tau)$ to the right of τ_0 is sufficiently pronounced, and may be much higher than τ_0 if the upturn or $T(\tau)$ is little pronounced.

I now explain, first informally and then more formally, why $T(\tau)$ is typically U-shaped. As when justifying concave increase of $R(\tau)$, the argument will be plausible but not fully transparent; again, more transparency might be achieved by considering a different decomposition of F, for instance that of Section 1.6 in which U-shaped toughness effects on the inclination parameters a_{nf} and a_{fn} may be easier to justify.

Informal justification of U-shapedness of $T(\tau)$.

- The policy of minimal toughness $\tau = \tau_{\text{min}}$ (that is, depending on the specific choice problem, detention of captured terrorists for a *zero* length of time, or absence of military presence, or absence of airport controls, etc.) entails a high number of terrorists $T(\tau) = T(\tau_{\text{min}})$ because of the absence of deterrence.

- A small increase of toughness τ beyond τ_{min} should reduce the number of terrorists $T(\tau)$, because the starting deterrence should overcompensate the starting hate effect; in other words, $T(\tau)$ is a decreasing function in some right neighbourhood of $\tau = \tau_{\text{min}}$. Indeed, a small amount of toughness (detention for a short length of time, or small military presence, or few airport controls, etc.), seems likely to create less hate than deterrence, by being perceived mainly as a legitimate amount of self-protection.

- Arguably, at least for the case of capture (rather than punishment) policy choice, if toughness τ exceeds some critical level (where this level is strongly case-specific),

29) of $F(\tau)$ is smaller than that of $T(\tau)$. The reason is that the growth rate of $F(\tau)$ equals that of $T(\tau)$ plus the growth rate of $1 - R(\tau)$, the latter being negative:

$$\frac{d}{d\tau}\log F(\tau) = \frac{d}{d\tau}[\log T(\tau) + \log(1 - R(\tau))] = \frac{d}{d\tau}\log T(\tau) + \frac{d}{d\tau}\log(1 - R(\tau)) < \frac{d}{d\tau}\log T(\tau).$$

additional toughness creates more additional hate effect than deterrence, thus letting $T(\tau)$ increase overall. Why this "taking over" of the hate effect? For instance, when choosing (some part of) capture policy, if toughness is already so high that a terrorist is more likely to be captured than to stay free, a further raise of toughness should provide little additional deterrence, while still providing additional hate. Here, deterrence has little growth potential once toughness is already high; by contrast, raising toughness from "moderate" to "high", or from "high" to "very high" may well significantly increase the hate effect, by being perceived *by the relevant group* as illegitimate.

More formal justification of U-shapedness of $T(\tau)$.

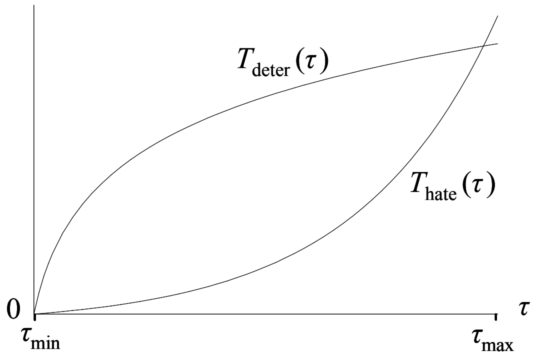

Figure 4: Examples of deterrence $T_{\text{deter}}(\tau)$ and hate effect $T_{\text{hate}}(\tau)$

The informal argument discussed above can be put more formally. If deterrence and hate have *linear* effects on T, then I may write $T(\tau) = T_0 - T_{\text{deter}}(\tau) + T_{\text{hate}}(\tau)$, where $T_{\text{deter}}(\tau)$ and $T_{\text{hate}}(\tau)$ are the deterrence effect resp. hate effect of toughness τ on T, and T_0 the value of $T(\tau)$ at toughness $\tau = \tau_{\text{min}}$ at which $T_{\text{deter}}(\tau)$ and $T_{\text{hate}}(\tau)$ are both zero. (In the quite possible case of non-linearity, a preceding transformation of T often leads to a linear model.[27]) As illustrated by Figure 4, the idea is that, typically,

(i) both deterrence $T_{\text{deter}}(\tau)$ and hate effect $T_{\text{hate}}(\tau)$ are strictly increasing functions,

(ii) deterrence $T_{\text{deter}}(\tau)$ is a strictly concave function (diminishing marginal hate effect),

(iii) hate effect $T_{\text{hate}}(\tau)$ is a strictly convex function (increasing marginal hate effect).

Before commenting on these claims, let us note what they entail. Supposing that first and second derivatives exist, (i) means that $T'_{\text{deter}}(\tau), T'_{\text{hate}}(\tau) > 0$, (ii) means that

[27] For instance, $T(\tau)$ may be the *product* (not sum) of a term depending only on deterrence and a term depending only on hate. This can be reduced to a linear (i.e. additive) relation by taking logarithms on both sides. In general, the (after transformation linear) model has the form $\tilde{T}(\tau) = T_0 - T_{\text{deter}}(\tau) + T_{\text{hate}}(\tau)$ where $\tilde{T}(\tau)$ is the transformation of $T(\tau)$ creating the linearity (e.g. $\tilde{T} = \log T$, or $\tilde{T} = \log \frac{T}{1-T}$, or any other increasing function of T). One could similarly argue that $\tilde{T}(\tau)$ is U-shaped (i.e. first decreasing, then increasing), which implies that $T(\tau)$ is U-shaped too and has the same minimising value as $\tilde{T}(\tau)$.

$T''_{\text{deter}}(\tau) < 0$, and (iii) means that $T''_{\text{hate}}(\tau) > 0$. Further, (ii)&(iii) imply that $T(\tau)$ is strictly convex, since $T''(\tau) = -T''_{\text{deter}}(\tau) + T''_{\text{hate}}(\tau) > 0$, and so we are already close to U-shapedness. If there exists a toughness $\tau_0 \in (\tau_{\min}, \tau_{\max})$ at which deterrence and hate effect have identical marginal returns ($T'_{\text{deter}}(\tau_0) = T'_{\text{hate}}(\tau_0)$), then this toughness τ_0 represents the unique minimum of $T(\tau)$, by $T'(\tau_0) = -T'_{\text{deter}}(\tau_0) + T'_{\text{hate}}(\tau_0) = 0$ and the strict convexity of $T(\tau)$. So, $T(\tau)$ would indeed be U-shaped, in fact even convex U-shaped.

Claim (i) is non-controversial: more toughness entails more deterrence and hate effects.

Claim (ii) seems plausible in many cases. Indeed, the rise of the deterrence effect $T_{\text{deter}}(\tau)$ in response to a marginal increase of toughness $\Delta\tau$ should be the lower, the higher τ already is: $T_{\text{deter}}(\tau)$ increases a lot if τ rises from 0 to 0.1, less if τ rises from 0.1 to 0.2, even less if τ rises from 0.2 to 0.3, etc. In the case of capture policy choice, once τ is so high that terrorists are more likely to be captured than to stay free, an additional rise of τ might leave deterrence $T_{\text{deter}}(\tau)$ nearly unchanged.

Claim (iii) is less obvious, and I do not wish to defend it for all situations. Whether (iii) holds depends on the psychological and sociological "dynamics of hate" (and similar feelings) in response to an increase in toughness. Only a case-by-case analysis can decide when (iii) holds, which is beyond the scope of this paper. A defence of (iii) would have to take the form of arguing that, the tougher the policy (already) is, the more the ground is prepared for an additional bit of toughness $\Delta\tau$ to create additional hate; for instance, an additional policeman ($\Delta\tau$) is the greater a provocation, the more policemen are already around, so the argument. Metaphorically, the additional policeman is the straw that breaks the camel's back *only if* there are already many policemen around. If this is the right metaphor, then additional toughness $\Delta\tau$ has more marginal hate effect if toughness τ is already close to what is perceived as provocative (resp. if the weight on the camel is already close to what the camel can maximally carry).[28]

But note that, even if (iii) fails to hold and $T_{\text{hate}}(\tau)$ is not strictly convex, $T(\tau)$ may still be strictly convex since $T''(\tau) > 0$ holds as soon as $T''_{\text{hate}}(\tau)$ exceeds the (negative!) quantity $T''_{\text{deter}}(\tau)$ for all τ. And even if the latter condition still fails and hence $T(\tau)$ is not strictly convex, $T(\tau)$ may still be U-shaped, i.e. first falling and then rising.

Exceptions. In some cases the function $T(\tau)$ is not convex U-shaped (see footnote 28), or not even U-shaped. Let me mention policy choice problems in which it is doubtful, respectively, (a) whether $T(\tau)$ has a decreasing part, or (b) whether $T(\tau)$ has an increasing part.

[28] Taking the metaphor seriously, if toughness τ is even higher, i.e. if the number of policemen is already perceived as provocative (resp. if the camel's back is already broken), then the additional police man $\Delta\tau$ may now bring little additional hate effect (it is impossible to do more than breaking the camel's back!). This would suggest that, when toughness τ grows from minimally tough $\tau_{\min}$ to maximally tough $\tau_{\max}$, the hate effect grows first slowly, then sharply (when moving from non-provocative to provocative), then again slowly; formally, hate effect $T_{\text{hate}}(\tau)$ would not be a convex function (as assumed in (iii)), but a first-convex-then-concave function. Under this scenario, $T(\tau) = T_0 - T_{\text{deter}}(\tau) + T_{\text{hate}}(\tau)$ would not be convex U-shaped, but of the shape of a "U" that stabilises (or even drops) at the top end of the toughness interval $[\tau_{\min}, \tau_{\max}]$. Provided that this stabilisation (or even drop) is not too pronounced, the optimal toughness or minimum of $F(\tau) = (1 - R(\tau))T(\tau)$ will still lay below the sharp rise of the hate effect $T_{\text{hate}}(\tau)$.

(a) Assume the group consists of persons who cannot be deterred at all, as it might seem to be the case for potential suicide bombers. (Note, however, that this is an extreme assumption since not only the actual terrorists, but also all non-terrorists would have to be entirely fearless.) Then toughness creates no deterrence ($T_{\text{deter}}(\tau) = 0$), and the remaining hate effect turns $T(\tau)$ $(= T_0 + T_{\text{hate}}(\tau))$ into an *everywhere increasing* function. More realistically, if deterrence is existent but small, it is doubtful whether $T(\tau)$ has a decreasing part since the marginal deterrence might be smaller than the marginal hate effect even for small toughness τ.

(b) In many policy choice problems related to punishment policy, it seems unclear whether $T(\tau)$ has an increasing part, i.e. whether marginal hate effect becomes higher than marginal deterrence for large toughness τ. One has to distinguish between choosing *length* and *conditions* of detention. In the first case (with toughness τ defined, say, as average number of years detained) there is little reason to assume an increasing marginal hate effect, i.e. to assume a convex hate effect $T_{\text{hate}}(\tau)$; perhaps even that $T_{\text{hate}}(\tau)$ flattens out for large toughness τ: An additional year of detention might be "a big deal" if added to 5 years of detention (much additional hate effect), but "no big deal" if added to 25 years of detention (little additional hate effect). Given such limited growth of $T_{\text{hate}}(\tau)$, it may be that $T_{\text{hate}}(\tau)$ never "takes over" compared to $T_{\text{deter}}(\tau)$, so that $T(\tau) = T_0 - T_{\text{deter}}(\tau) + T_{\text{hate}}(\tau)$ stays decreasing even for large τ. As for choosing conditions of detention (size of cells, number of hours of work per day per prisoner, etc.), there is here no general reason to assume decreasing marginal deterrence, i.e. concavity of $T_{\text{deter}}(\tau)$. For instance, when choosing the number of hours work per day, $T_{\text{deter}}(\tau)$ may be convex rather than concave: One additional hour is bearable when added to 5 hours work (little additional deterrence), but perhaps unbearable when coming on top of 10 hours work (much additional deterrence). So, as when choosing length of detention, the hate effect need not "take over" and hence $T(\tau)$ need not be increasing for large τ, but this time because $T_{\text{deter}}(\tau)$ is not concave (perhaps convex) rather than because $T_{\text{hate}}(\tau)$ is not convex (perhaps concave). In fact, when choosing conditions of detention the shape of $T(\tau)$ may even be unclear everywhere since even the increasing start of $T(\tau)$ is questionable. Indeed, it does not seem obvious whether an initial raise of toughness of conditions of detention (one hour of work instead of no work, etc.) creates more deterrence or more hate effect on $T(\tau)$.

2.5 Summary of typical shapes and general policy recommendations

I have argued that $T(\tau)$ is typically U-shaped and $R(\tau)$ typically concave increasing, with important exceptions in the field of punishment policy choice, i.e. of choosing length and conditions of detention. The typical shapes are summarised in Table 1, where "unclear" stands for "no typical shape". Further, "capture policy" stands for a policy choice regarding *either* the entire capture policy *or* of one of its components (such as number of airport controls); and similarly for "conditions of detention". Also, recall that length of detention and conditions of detention together make up punishment policy, and that punishment policy and capture policy together make up the symptom-related policy.

In Table 1, the entries for the function $R(\tau)$ were justified in Section 2.3 (with conditions of detention mentioned under "exceptions"), and the entries for the func-

	function $R(\tau)$	function $T(\tau)$	optimal toughness level τ
capture policy	increasing & concave	U-shaped	$<$ where MH gets significant $>$ where MH equals MD
length of detention	increasing & concave	first decreasing then unclear	not very low
conditions of detention	flat	unclear	unclear

Table 1: Typical shapes of $R(\tau)$ and $T(\tau)$ and policy recommendations for different policy choice problems (MH: marginal hate effect; MD: marginal deterrence)

tion $T(\tau)$ were justified in Section 2.4 (with length of detention and conditions of detention mentioned under "exceptions"). The policy recommendations in the last column are, of course, no more than rough indications to be refined in case-by-case analyses. The recommendations follow from the two preceding columns together with the assumption that the (sole) aim is to minimise number of free terrorists $F(\tau) = (1 - R(\tau))T(\tau)$. The policy recommendation for capture policy has been discussed in more detail in Section 2.4 under "optimal toughness". Regarding length of detention, toughness should be "not very low" because the initial decrease of $T(\tau)$ and the increase of $R(\tau)$ imply that $F(\tau)$ has an initial decrease, and hence toughness should be beyond this region of initial decrease. No general policy recommendation is possible for conditions of detention, given the absence of typical properties of the function $F(\tau)$.

2.6 Two extreme cases: flat $T(\tau)$ and V-shaped $T(\tau)$

It is worth highlighting the implications of a U-shaped function $T(\tau)$ – the typical case at least for capture policy choice – by considering two extreme (but often unrealistic) scenarios. These scenarios might be informally described as the case where the "U" is (close to) a flat line and the case where the "U" is (close to) a "V". More precisely, in the first case the growth rate of $T(\tau)$ stays near zero, and in the second case the growth rate of $T(\tau)$ rises sharply from "very negative" to "very positive".[29]

The function $T(\tau)$ is flat (Figure 5) if people react only very little to the toughness level, so that the number of terrorists is roughly the same whatever the toughness. Then, minimising $F(\tau) = (1 - R(\tau))T(\tau)$ is nearly equivalent with minimising the first factor $1 - R(\tau)$, i.e. with maximising detention rate $R(\tau)$. So, since $R(\tau)$ is a decreasing function, the policy should be maximally tough (in my model which neglects all costs of toughness).

By contrast, in the case of a V-shaped function $T(\tau)$ (Figure 6), in which the number of terrorists is highly sensitive to the toughness level, minimising $F(\tau) = (1 - R(\tau))T(\tau)$ is nearly equivalent with minimising the second factor $T(\tau)$.[30] So,

[29] The growth rate of a (positive and differentiable) function $g(\tau)$ is defined as $\frac{d}{d\tau}\log g(\tau) = g'(\tau)/g(\tau)$. This measures the "relative growth", while the derivative $g'(\tau)$ measures the "absolute growth".

[30] Let me be more precise without over-formalising. The claim that $F(\tau)$ and $T(\tau)$ reach their minima for roughly the same toughness τ means that the derivatives $F'(\tau)$ and $T'(\tau)$ are zero for roughly the same τ, or that the growth rates $\frac{d}{d\tau}\log F(\tau)$ and $\frac{d}{d\tau}\log T(\tau)$ are zero for roughly the same τ. By the additivity of growth rates, the growth rate of $F(\tau)$ is the sum of the growth rates of

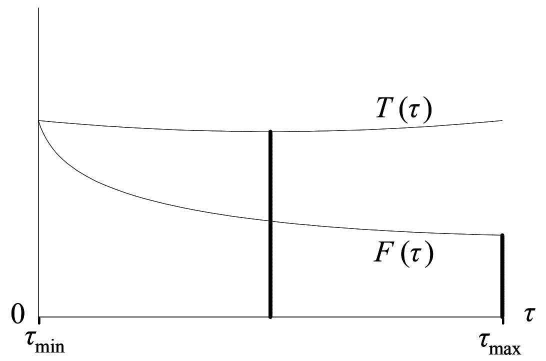

Figure 5: Example of flat function $T(\tau)$ and resulting function $F(\tau) = (1-R(\tau))T(\tau)$; minima of $T(\tau)$ and $R(\tau)$ indicated by thick lines

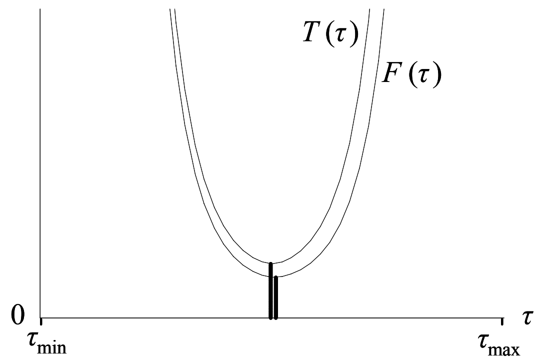

Figure 6: Example of V-shaped function $T(\tau)$ and corresponding function $F(\tau) = (1-R(\tau))T(\tau)$; minima indicated by thick lines

the policy should aim to minimise the number of terrorists regardless of the capture success, i.e. to implement the particular toughness level at which the combination of deterrence and hate effect entails the minimal inclination towards terrorism within the group (marginal hate effect equal to marginal deterrence).

It might seem that the policy of the "international anti-terrorism coalition" in response to September 11, 2001, is mainly concerned with increasing the detention rate R, thereby implicitly assuming the case of a flat function $T(\tau)$.

the two factors $1-R(\tau)$ and $T(\tau)$ (indeed, $\frac{d}{d\tau}\log F(\tau) = \frac{d}{d\tau}\log[(1-R(\tau))T(\tau)] = \frac{d}{d\tau}\log(1-R(\tau)) + \frac{d}{d\tau}\log T(\tau)$). But the V-shapedness of $T(\tau)$ means by definition that its growth rate $\frac{d}{d\tau}\log T(\tau)$ rises sharply from "very negative" to "very positive" (in fact, I only need $\frac{d}{d\tau}\log T(\tau)$ to rise sharply once it becomes positive). So, adding to $\frac{d}{d\tau}\log T(\tau)$ the function $\frac{d}{d\tau}\log(1-R(\tau))$ (which stays much closer to 0) will lead to a function $\frac{d}{d\tau}\log F(\tau)$ that "crosses the τ-axis" at roughly the same point as $\frac{d}{d\tau}\log T(\tau)$ does. In other words, $\frac{d}{d\tau}\log F(\tau)$ and $\frac{d}{d\tau}\log T(\tau)$ are zero for roughly the same toughness levels τ, as claimed.

3 A methodological assessment of the approach

I have illustrated how the general method (a)-(d) – decomposing an objective function into inclination and neutralisation parameters and estimating policy effects on each of these parameters – may work in the case of decomposing the number of free terrorists F into $(1-R)T$. The estimation step has been theory-based rather than statistical, and may be refined in a case-by-case analysis. While the last section dealt exclusively with the symptom-related policy and its toughness level, I now go back to the general discussion, while sticking to the same decomposition as an illustration. Since I consider an arbitrary, more or less narrow (anti-terrorism) policy choice problem, the term "policy" refers to *either* the full anti-terrorism policy, *or* some relevant component of it, such as (part of) the cause-related policy, or (part of) the symptom-related policy resp. the toughness level. A policy in this flexible sense is denoted by π, and the set of policies considered by $\boldsymbol{\pi}$. Examples include the choice of the amount of development aid for a given country (cause-related), the decision whether to attack a given country, and the choice of length or conditions of detention (symptom-related). The reader who prefers to focus on a particular type of policy choice may replace in his/her mind π either by (χ, σ) (choosing the full anti-terrorism policy) or by χ (choosing (some part of) the cause-related policy) or by σ resp. τ (choosing (some part of) the symptom-related policy resp. its toughness level).

But is the "indirect" method (a)-(d) adequate, or would it have been better to focus directly on F and forget about the decomposition into $(1-R)T$? To answer this question, in this section the so-called *direct statistical* estimation (of policy effects) is compared with the *theory-based* estimation (first steps of which were made in this paper). While the former is purely statistical, the latter is based on theoretical arguments, which are ideally empirically/statistically tested. It is argued, perhaps not as a final word, that a direct statistical estimation – if possible at all – does *not* benefit from the decomposition into $(1-R)T$ (but perhaps from other decompositions), whereas a theory-based estimation – often the only possibility – should indeed decompose F (for instance into $(1-R)T$), to cut a complex estimation problem into more manageable pieces. After clarifying the difference between direct statistical and theory-based estimation in Section 3.1, I discuss the purely statistical estimation in Sections 3.2 and 3.3, and the theory-based estimation in Sections 3.4 and 3.5.

3.1 Clarifying the terminology

After some general words about "theoretic" versus "statistical" arguments, I will make the distinction between "direct statistical" and "theory-based" estimation. The former is purely statistical, the latter based on (ideally, statistically tested) arguments.

"Theoretic" versus "statistical". In politics and economics it is common to distinguish between theoretical and statistical arguments or analyses. As a simple example, consider the *demand function* and the *supply function* for a given good in some market economy. Straightforward reasoning about the aims of demanders and suppliers tells (at least) that demand is a decreasing and supply an increasing function of price – a simple example of theoretical arguments. A statistical analysis of past price and demand/supply data may allow a more precise estimation of both curves (after overcoming high obstacles since prices are endogenous). More generally, when trying to

learn how different variables of interest affect each other (how prices affect demand and supply, how interest rates affect inflation, how climate affects birth rate, and in our case how policies π affect the values of F, T, R), a theoretical analysis proceeds by understanding the relevant *causal mechanisms* governing the interaction between variables, using psychological, sociological, political, economical, or indeed any type of theoretical arguments. As in this paper, this results typically (but not always[31]) in *qualitative* conclusions: increasing or decreasing effect (first derivative), increasing or decreasing marginal effect (second derivative), existence of minima or maxima, etc. By contrast, a statistical analysis is based on past data and may, in principle, be "blind" in the sense of not caring of why and by what means the estimated effects occur. Of course, theoretical and statistical arguments may be combined.[32]

"Theory-based" and "direct statistical" estimation. The effect of policies π on values of F, T, R, as given by the functions $F(\pi), T(\pi), R(\pi)$ ($\pi \in \boldsymbol{\pi}$), are somewhat more difficult effect to estimate (theoretically or statistically!) than, say, the effect of prices on demand or supply. What I want to compare are not theoretical and statistical arguments *per se*, but the "direct statistical" and the "theory-based" estimation of the functions $F(\pi)$, $T(\pi)$, and $R(\pi)$. With a "direct statistical" estimation of $F(\pi)$ (resp. $T(\pi), R(\pi)$) I mean a statistical estimation based on data of how policies π affected F (resp. T, R) in the past. A "theory-based" estimation of (mainly qualitative features of) the function $F(\pi)$ (resp. $T(\pi), R(\pi)$) need not be "purely" theoretical: It is based on arguments that have ideally been verified empirically/statistically and in this sense may be a combined theoretical and statistical approach; the difference to the *direct* statistical estimation is that the statistical data used (if any) are not data directly about policy effects on F (resp. T, R) but other data that confirm certain hypotheses used in an (otherwise theoretical) estimation of policy effects on F (resp. T, R). For instance, when estimating how the policy π of an improved school system affects the number of terrorists T, a direct statistical estimation would use past data of how T responded to improved education, while a theory-based estimation might use statistically confirmed arguments about psychological, sociological, or other effects of higher levels of education, such as effects on the "youth culture", on group dynamics, on the aims and wishes of young people, on the crime rate among young people, and so on.

3.2 Direct statistical estimation: difficult for each of F, T, R

Let me briefly summarise the rather serious obstacles faced by a direct statistical estimation of the functions $F(\pi), T(\pi), R(\pi)$, i.e. of the causal effects of policies on values of F, T, R. I argue that three standard econometric problems are at their strongest: the lack of data, the difficulty to measure data, and a lack of ceteris paribus. (A partial escape-route as far as F is concerned is discussed in the next subsection.)

Unprecedented policy choice. If the policy choice problem faced is new and many or all of the policies $\pi \in \boldsymbol{\pi}$ (or approximations of them) have never been in action

[31] A counterexample is *game theory*, which derives the *exact* behaviour of (rational) agents from theoretical arguments about rationality.

[32] In fact, an analysis is rarely *purely* statistical, since the way data are used often requires at least some theoretical knowledge about, for instance, which regressors to include into a regression equation and which of them to treat as exogenous.

in the past, then the direct statistical estimation fails simply because of the absence of past data. This might seem to be the case for the capture policy choice faced by the international anti-terrorism coalition after September 11, 2001. But other capture policy choices on smaller scales need not be unprecedented, nor are many punishment policy choices and many cause-related policy choices such as the kind and extent of development policy.

Difficulty to measure F, T, R. Assume that a given policy π (or a close approximation of it) has been run in the past (for a sufficiently long period of time and relative to some group preferably similar to the one presently relevant). Can the resulting values of F, T, R be measured? What is easily measured is the number of detained terrorists D, but it is unclear how the rest of the group is divided up into a number of non-terrorists N and free terrorists F since free terrorists will try not to appear as such. F can at most be estimated, for instance through secret service information. Knowing D but not F and N, one knows neither of the parameters of interest: number of free terrorists F, number of terrorists $T = F + D$, and detention rate $R = D/T = D/(F + D)$ – a serious problem for "data collection".

Lack of ceteris paribus. By ignoring the two problems just mentioned, assume that a given policy $\pi \in \boldsymbol{\pi}$ was in action in the past, say exactly one time, and that the resulting values of F, T, R were measured; let F', T', R' be the values measured. Let us see why these values may be of little indication for the present choice situation. Following common practice in statistics, the estimate of a parameter is written with a "hat". Having exactly one past observation (sample size of one), it is natural to estimate the true values $F(\pi), T(\pi), R(\pi)$ by the following group-size-adjusted version of the past-observed values:

$$\hat{F}(\pi) := \frac{m}{m'}F',\ \hat{T}(\pi) := \frac{m}{m'}T',\ \hat{R}(\pi) := R', \tag{4}$$

where m and m' denote the size of the present and the past-observed group, respectively. For instance, if the present group is twice as large as the past-observed group ($m/m' = 2$), then $F(\pi)$ is estimated by $\hat{F}(\pi) = 2F'$, i.e. by twice the number of free terrorists measured in the past group. Note that I need no group-size-adjustment to estimate detention rate $R(\pi)$ since $R(\pi)$ is a *ratio* of absolute numbers. Can $\hat{F}(\pi), \hat{T}(\pi), \hat{R}(\pi)$ be expected to be good approximations of the true $F(\pi), T(\pi), R(\pi)$? Usually no, because F, T, R are influenced by factors other than the policy π, and these other factors will usually vary between the past and the present group (no ceteris paribus): different level of poverty, level of education, age structure, group cohesion, climate, etc. Basic econometrics tells us that estimates such as (4) can be relied upon only if factors other than the policy *either* have a negligible impact on F, T, R *or* happen to be approximately the same for the past and the present case (ceteris paribus). Unfortunately, none of these conditions is usually met.[33] For instance, the effects of a given policy π applied to the Afghan population, while perhaps reliably estimated by the effects of the same policy applied to the same population two years earlier (approximate ceteris paribus), cannot be estimated by

[33] A drastic example might illustrate this: The fact that the (population-size-adjusted) values of T and F for the French population are easily ten, hundred or more times lower than for the Palestinian population is caused not only by different anti-terrorism policies, but also to factors such as more social stability, less poverty and higher education in France.

the effects of the same policy applied to some other population living in different circumstances (no ceteris paribus). So, estimates such as (4) have to be treated with great caution, except in the rare cases of (approximate) ceteris paribus.

If the effect of policy π has been measured k times rather than a single time, there is a hope of obtaining more reliable estimates of $F(\pi), T(\pi), R(\pi)$, provided that the different violations of ceteris paribus across the sample roughly cancel out against each other, that is, behave *in average* like in the present situation rather than deviating systematically from it.[34] However, this "cancelling out" is not only unlikely, but also hard to define and verify, so that we are far from exact statistics.

Of course, more sophisticated estimation techniques than that of (4) resp. footnote 34 may be available, such as a (linear or non-linear) regression analysis.[35] All techniques will in one way or another face the problem of lack of ceteris paribus; but some techniques may be more immune than others by tackling the problem explicitly (e.g. the inclusion of control variables in a regression, or a 2-stage regression).

3.3 Direct statistical estimation: easier for F than for T and R

Let us now see why, luckily, a policy choice is possible without knowledge of the policy effects on *absolute levels* of F, T, R – knowledge of certain *relative changes* is sufficient; and, so my argument will continue, direct statistical estimation of relative changes is easier for F than for T and R, because the measurement problem may be overcome for F. This suggests that a direct statistical estimation – if possible at all – should focus straight on F and ignore the explanation in terms of T and R. (I leave open whether this argument can be extended to other decompositions of F.)

To simplify, let us assume that the choice is between only two policies, one of which might be the status quo and the other an alternative. The two policies are denoted 1 and 2, with corresponding values of F, T, R denoted F_1, T_1, R_1 resp. F_2, T_2, R_2. The question of which policy is better can be expressed either directly in terms of F or indirectly in terms T and R:

	in terms of F	in terms of T, R
policy 2 is better than (is worse than, ties) policy1 if...	$\frac{F_2}{F_1} > (<, =) 1$	$\frac{T_2}{T_1} > (<, =) \frac{1-R_1}{1-R_2}$.

Table 2: Decision criterion based on F and based on T, R

[34] Measuring F, T, R in k different situations results in a series of triples $(F^1, T^1, R^1), ..., (F^k, T^k, R^k)$. Based on this sample of size k, the natural estimates of $F(\pi), T(\pi), R(\pi)$ are given by taking the average (population-size-adjusted) observations:

$$\hat{F}(\pi) := \frac{1}{k}\left[\frac{m}{m^1}F^1 + ... + \frac{m}{m^k}F^k\right],\ \hat{T}(\pi) := \frac{1}{k}\left[\frac{m}{m^1}T^1 + ... + \frac{m}{m^k}T^k\right],\ \hat{R}(\pi) := \frac{1}{k}\left[R^1 + ... + R^k\right],$$

where $m^1, ..., m^k$ are the group sizes in the past observations, respectively.

[35] The estimates $\hat{F}(\pi), \hat{T}(\pi), \hat{R}(\pi)$ of $F(\pi), T(\pi), R(\pi)$ for a given policy π need not be based solely on those values of F, T, R observed for that policy π, but may be based more generally on all values of F, T, R observed for policies π' in $\boldsymbol{\pi}$ (or even outside $\boldsymbol{\pi}$). This is the case in a regression analysis. A (linear or non-linear) regression of F, T and R may become possible if each policy π can be identified with a real number (e.g. a toughness τ or an amount of money spent for development aid) or with a vector of real numbers (e.g. a multi-dimensional toughness parameter $(\tau_1, ..., \tau_k)$ or a vector of amounts spent for different issues).

(The second criterion follows from $F_i = (1 - T_i)R_i$.) Whichever criterion is used, knowledge of absolute magnitudes of F resp. T, R is not required; rather, knowledge of the *relative change* of F resp. of T and $1 - R$ is sufficient:

(a) Using the criterion based on F, an estimate of the ratio F_2/F_1, and more precisely of how F_2/F_1 compares to 1, suffices; no idea of the absolute magnitudes F_1 and F_2 is required.

(b) Using the criterion based on T, R, estimates of the ratios T_2/T_1 and $(1 - R_1)/(1 - R_2)$, and more precisely of how they compare to each other, suffice; no idea of the absolute magnitudes T_1, T_2, R_1, R_2 is required.

Which ratios are easier to estimate: that in (a) based on T or those in (b) based on T, R? Given that the estimation is *direct statistical*, it is that in (a), as is seen now (whereas in a *theory-based* estimation it is those in (b), as seen later). Let me explain. In (a) one has to estimate the ratio $F_{12} := F_2/F_1$, and in (b) the ratios $T_{12} := T_2/T_1$ and $R_{12} := (1-R_1)/(1-R_2)$. Estimating ratios rather than absolute magnitudes does usually not help avoiding the first and third of the problems mentioned in Section 3.2 (unprecedented policy choice and lack of ceteris paribus). However, the measurement problem may be overcome for F_{12}, but not so for T_{12} and R_{12}. The "trick" is that, while number of free terrorists cannot be measured, this number can, plausibly, be assumed as roughly proportional to the number of terrorism attacks which *can* be measured. For policies 1 and 2 respectively, denote by A_1 and A_2 the resulting *attack frequency*, i.e. the number of attacks per unit of time (month, or quarter, or year, etc.). Assuming that attack frequency is roughly proportional to number of free terrorists, $F_2/F_1 \approx A_2/A_1$.[36] For instance, if there are twice as many attacks per unit of time under policy 2 than under policy 1 ($A_2/A_1 = 2$) then there are roughly twice as many free terrorists under policy 2 than under policy 1 $F_2/F_1 \approx 2$). It follows that an estimate of the ratio $A_{12} := A_2/A_1$ can serve as an estimate of $F_{12} = F_2/F_1$.[37] Whether A_1 and A_2 and hence A_{12} may indeed be estimated from past-observed attack frequencies depends on whether the two other obstacles exist or may be overcome (unprecedented policy choice, and lack of ceteris paribus).[38]

3.4 Theory-based estimation: how to go beyond my analysis

Let me now turn to the theory-based estimation of policy effects on F, T, R, which does not use past data of F, T, R but is based on theoretical arguments (which may still be empirically/statistically confirmed). I start, in this subsection, by giving two reasons of why a theory-based estimation may be successful and go beyond the first steps made in this paper.

More results by specialising. By focussing on a concrete policy choice problem rather than taking the general angle of this paper, more refined and concrete estima-

[36] Insofar as this approximation is inaccurate, the policy will minimise the attack frequency rather than the number of free terrorists – perhaps no big problem.

[37] Although such an estimation is based on data of policy effects on attack frequencies, not on F itself, we may still talk of "direct statistical estimation" (in a wider sense) because data of (relative changes of) attack frequencies may count as approximate data of (relative changes of) F.

[38] The simplest estimate of A_1 resp. A_2 is the average (group-size-adjusted) past-observed attack frequency under policy 1 resp. 2 (as for the estimates of $F(\pi)$ and $T(\pi)$ in footnote 34). But a regression analysis, perhaps with control variables, may also be possible, perhaps even advisable in order to compensate for a lack of ceteris paribus.

tion of policy effects on F, T, R may become possible. For instance, the statement that the number of terrorists T is typically a U-shaped function of toughness could be refined by estimating the *location* of the minimum and perhaps the *extent* of the "upturn" after the minimum, once one considers a specific capture policy problem with regard to a specific group with a particular sociological, historical and economical background. For example, which level of toughness against the terror organisation Al Quaida minimises the number of new terrorists (perhaps new Al Quaida members) in the Middle East? And if this level of toughness is exceeded, how strongly does T rise? This is a question of the deterrence and hate effect of toughness specifically for the population of the Middle East.

Knowing relative changes suffices. Like a direct statistical estimation, a theory-based estimation of *absolute* magnitudes may be harder than one of *relative* changes. Luckily, as discussed in the previous subsection, knowledge of relative changes suffices for policy optimisation, whether one estimates F directly or goes for the decomposition and estimates policy effects on T and R. In a choice between only two policies 1 and 2, one may use one of the criteria of Table 2, which are based on *ratios* rather than absolute magnitudes. More generally, in order to minimise $F(\pi) = (1 - R(\pi))T(\pi)$ it is sufficient to estimate *up to a (positive) multiplicative constant* the function $F(\pi)$ (resp. the functions $T(\pi)$ and $1 - R(\pi)$ if one goes for the decomposition).[39] Such estimates are obtained for instance by, *for every policy* $\pi \in \boldsymbol{\pi}$, estimating the relative change of F (resp. T and $1 - R$) when the status quo policy is replaced by policy π, that is, answering the question "By which percentage would F (resp. T and $1 - R$) rise or fall if policy π were implemented?"; absolute changes of F (resp. T and $1 - R$) need not be estimated and can stay completely unknown.[40,41]

3.5 Theory-based estimation: the need to decompose F

As argued earlier, a *direct statistical* estimation of policy effects – if possible at all – does not benefit from decomposing F into $(1 - R)T$. By contrast, I now argue that a theory-based estimation should derive policy effects on F via the "indirect" strategy of estimating policy effects on the inclination and neutralisation parameters in some suitable decomposition of F, for instance that into $(1 - R)T$ or that given in Section 1.6. This defends the method (a)-(d) as being right in a theory-based context.

[39] Knowing the functions $F(\pi)$ (and similarly for the functions $T(\pi)$ and $1 - R(\pi)$) *up to a multiplicative constant* means knowing that $F(\pi) = a \times \widetilde{F}(\pi)$ for all $\pi \in \boldsymbol{\pi}$, with known function $\widetilde{F}(\pi)$ but unknown constants a (> 0), hence known ratios $F(\pi_2)/F(\pi_1) = \widetilde{F}(\pi_2)/\widetilde{F}(\pi_1)$ $(\pi_1, \pi_i \in \boldsymbol{\pi})$ since a cancels out.

[40] The answers to these questions yields the ratio $F(\pi)/F_0$ (resp. $T(\pi)/T_0$ and $(1-R(\pi))/(1-R_0)$) where F_0, T_0, R_0 are the parameters in the status quo; and so we know $F(\pi)$ (resp. $T(\pi)/T_0$ and $1 - R(\pi)$) up to a multiplicative constant, namely up to the unknown factor $1/F_0$ (resp. $1/T_0$ or $1/(1 - R_0)$).

[41] In the case of choosing between cause-related policies, both strategies (directly estimating relative changes of F or going for the decomposition of F) are in fact formally, but not interpretationally, equivalent. To see why, note that each policy results in the same detention rate R, i.e. we have a constant function $R(\pi) = R_0$ (see Section 1.2). So, by $F(\pi) = (1 - R_0)T(\pi)$, the functions $F(\pi)$ and $T(\pi)$ differ by a multiplicative constant, and hence have identical *relative* changes. Despite this formal equivalence, the theory-based analysis should be *interpreted* as one of T, not one of F, since it consists in estimating policy effects on people's inclination towards terrorism, for which T stands whereas F stands for a certain combination of inclination (T) and a detention rate (R).

The aim is to estimate effects of a variable X (here: the policy π) on a variable Y (here: number of free terrorists F). By definition, a theory-based estimation proceeds by analysing the causal mechanisms by which X affects Y, i.e. by understanding the precise means through which this effect comes about, ideally based on empirically confirmed arguments. It is advisable to cut a complex problem into more tractable pieces, specifically into its semantically distinct components. Accordingly, the overall effect of X on Y is better theoretically understood by analysing the different subeffects or means by which X affects Y. In the case of effects of policies on F, this may be achieved by subdividing policies into components (the roughest subdivision being into cause- and symptom-related policy) and to decompose F into inclination and neutralisation parameters (e.g. T and R), thereby allowing separate estimation of effects of each policy component on each parameter. While the choice of the decomposition is flexible, the need for a decomposition seems evident from a theoretical perspective: F is the number of free terrorists, i.e. of persons defined by two properties of different nature and origin: "being a terrorist" and "being not yet captured (or already released)". Of these two properties, whether a person has the first is a question of his or her inclination (towards terrorism), and whether he or she has the second is a question of how much the symptom-related policy neutralises the terrorism threat. Specifically, inclination is measured in the present model by T, in the model of Section 1.6 by the frequencies a_{nf} and a_{fn} (of non-terrorists becoming free terrorists and vice versa), in even other models perhaps by different inclination parameters for each age group, etc. And neutralisation is measured in the present model by R, in the model of Section 1.6 by the frequency a_{fd} of capturing free terrorists, in even other models perhaps by the average time a terrorist stays free and the average time a captured terrorist is detained, etc. Policy effects on inclination parameters may be estimated using psychological, sociological, economical, cultural, religious, and other arguments. Policy effects on neutralisation parameters may be estimated using criminological, military, technological, infrastructural, geographical, and other arguments. Such estimations are difficult enough, but seemingly easier than at once estimating the combined policy effect on F.

4 Summary and conclusion

Let me give a summary of each section, and finish with some concluding remarks.

Section 1. To illustrate the method (a)-(d) for anti-terrorism policy choice, I have applied it to the simple objective function of minimising the number of free terrorists F (see Section 1.1 for the additional inclusion of costs). F was decomposed into the product $(1-R)T$, where the number of terrorists T reflects the group's inclination towards terrorism, and the detention rate R reflects the success in neutralising terrorists' power. Some alternative decompositions of F with perhaps more transparent inclination and neutralisation parameters than T and R were discussed in Section 1.6. In order to reduce $F=(1-R)T$, the cause-related policy reduces inclination (T), and the symptom-related policy increases neutralisation (R). But the symptom-related policy, which is composed of the capture policy and the punishment policy (and the defence policy, ignored here given that the objective function F neglects how powerful a free terrorist is), also has (side) effects on inclination (T), whose overall direction is not obvious since (desirable) deterrence and (undesirable) hate effects compete against each other.

Section 2. In this section, I have focussed on choosing (some part of) the symptom-related policy, and more precisely the "toughness level". Toughness was, for simplicity, measured by a unidimensional parameter τ, which, depending on the specific policy choice problem, may be size of an occupying army, number of airport security checks, length of detention, size of prison cells, etc. I have analysed the typical effects that toughness has on the number of terrorists T and the detention rate R, by distinguishing between capture policy choice, choice of length of detention, and choice of conditions of detention. Table 1 summarises the typical shapes of $R(\tau)$ and $T(\tau)$ and the policy recommendations suggested by these shapes. In short, capture policy should typically be not very tough, and length of detention typically not very short; no such recommendations were derived for conditions of detentions. Finally, I have considered two stereotypical (but probably not realistic) cases: If people's inclination towards terrorism reacts very little to the toughness level (flat function $T(\tau)$), minimising $F(\tau)$ is nearly equivalent with minimising the first factor $1 - R(\tau)$, i.e. maximising toughness τ; and if people's reaction to the toughness level is very strong (V-shaped function $T(\tau)$), minimising $F(\tau)$ is nearly equivalent with minimising the second factor $T(\tau)$, i.e. creating the climate in which people are least inclined towards terrorism, regardless of the capture success.

Section 3. I then turned to a methodological assessment of this strategy (a)-(d) of policy optimisation, again by means of the example of the objective function F. In order to estimate policy effects on F, is it justified to decompose F (e.g. into $(1 - R)T$) and estimate separately policy effects on each parameter of the decomposition (e.g. T, R)? The answer, I have argued, depends on the estimation technique. While a direct statistical estimation of policy effects on a variable (e.g. F or T or R) is based on statistical data of how the variable responded to policies in the past, a theory-based estimation is based on (ideally empirically/statistically tested) arguments about the causal mechanisms through which policies affect the variable – for instance psychological and sociological arguments for the variable T, and criminological and military arguments for the variable R. I have argued that a direct statistical estimation of policy effects on F or T or R is often impossible, because three standard econometric problems are at their strongest: unprecedented policy choice, problems in measuring F or T or R, and a lack of ceteris paribus between the past and present situation. However, if a direct statistical estimation is nevertheless attempted, it is probably advisable to estimate straight the effects on F, because the measurement problem may be overcome for F by using attack frequencies as indicators for the number of free terrorists. The question of whether other decompositions of F may be more helpful for a direct statistical estimation is left open. By contrast, if a theory-based estimation is chosen (often the only possibility), a decomposition of F (e.g. into $(1-R)T$) reduces a complex estimation problem to more manageable ones: estimating effects of the cause-related policy on inclination parameter(s), effects of the symptom-related policy on neutralisation parameter(s), and deterrence and hate effects of the symptom-related policy on inclination parameter(s). Such theory-based estimation may go beyond the first steps done in this paper by considering specific policy choice questions.

Concluding remarks. Let me finish with some considerations on the endogeneity of people's inclination towards terrorism, i.e. the fact that inclination may be influenced by policies. Perhaps the latter is easily overseen in reality, because it easier to track symptom-related policy success (captures, thwarted attacks, etc.) than effects on inclination. Indeed, who can ever prove what has caused an increase or decrease of the number of terrorists? But this does not imply that such effects are inexistent or insignificant. It might seem that the international anti-terrorism coalition after September 11, 2001 is running a mainly symptom-related policy. Is there here an implicit assumption that inclination is exogenous, i.e. not influenced and not ineffaceable by policies? If it is true that the number of terrorists T is exogenous, then there is neither a need for a cause-related policy, nor a need to care of effects of the symptom-related policy on inclination (flat function $T(\tau)$). Clearly, maximal toughness would then be appropriate (in my model which neglects all costs of toughness). But if inclination is endogenous, it may become crucial to also run a cause-related policy and to care of effects that the symptom-related policy has on inclination.

References

Bovens, L. and W. Rabinowicz (2001) Democratic Answers to Complex Questions - an Epistemic Perspective, *Synthese*, forthcoming.

Brophy-Baermann, R. and J. A. C. Conybeare (1994) Retaliating against terrorism: rational expectations and the optimality of rules versus discretion. *American, Journal of Political Science* 38 (1), p. 196-210.

Cioffi-Revilla, C. (1998) *Politics and uncertainty*, Cambridge University Press.

Cioffi-Revilla, C. (1998a) On the likely magnitude, extent, and duration of an Iraq-UN war, *Journal of Conflict Resolution* 35, p. 387-411.

Cioffi-Revilla, C. (1985) Political reliability theory and war in the International system, *American Journal of Political Science* 29, p. 47-68.

Cioffi-Revilla, C. and H. Starr (1995) Opportunity, willingness, and political uncertainty: theoretical foundations of politics, *Theoretical Journal of Politics* 7, p. 447-76.

Congleton, R. D. (2002) Terrorism, interest-group politics, and public policy, *The Independent Review* 7 (1).

Dietrich, F. (2004) Dynamic models of terrorism prevention, *working paper*, University of Konstanz.

Dumas, L. J. (2002) Is development an effective way to fight terrorism? *Philosophy & Public Policy Quarterly* 22 (4), p. 7-12.

Frey, B. S. and S. Luechinger (2003) How to fight terrorism: alternatives to deterrence. Defence and Peace Economics 14(4), p. 237-49.

Grofman, B. and S. L. Feld (1988) Rousseau's General Will: A Condorcetian Perspective, *American Political Science Review* 82 (2), p. 567-76.

Jones, B., B. Radcliff, Ch. Taber, R. Timpone (1995) Condorcet winners and the paradox of voting: probability calculations for weak preference orders, *American Political Science Review* 89, p. 113-145.

List, C. (2003) The Probability of Inconsistencies in Complex Collective Decisions, *Social Choice and Welfare*, forthcoming.

List, C. and R. E. Goodin (2001) Epistemic Democracy: Generalizing the Condorcet Jury Theorem, *Journal of Political Philosophy* 9 (3), p. 277-306.

Pape, R. A. (2003) The Strategic Logic of Suicide Terrorism, American Political Science Review 97 (3), p. 343-361.

Pettit, P. (2001) Deliberative Democracy and the discursive Dilemma, *Philosophical Issues* 11 (Supp. *Noûs*), p. 268-299.

Sandler, T., J. T. Tschirhart, and J. Cauley (1983) A Theoretical Analysis of Transnational Terrorism, American Political Science Review 77 (1), p. 36-54.

Sandler, T. and W. Enders (2002) An economic perspective on transnational terrorism. *Working paper*, University of Southern California.

Tangian, A. S. (2000) Unlikelihood of Condorcet's paradox in large societies, *Social Choice and Welfare* 17, p. 337-365.

Wulf, Wm. A., Y. Y. Haimes, and T. A. Longstaff (2003) Strategic alternative responses to risks of terrorism. *Risk Analysis* 23 (3), p. 429-44.