Utility Proportional Beliefs

Christian W. Bach and Andrés Perea*

Department of Quantitative Economics
School of Business and Economics
Maastricht University
6200 MD Maastricht
The Netherlands

c.bach@maastrichtuniversity.nl a.perea@maastrichtuniversity.nl

Abstract. In game theory, basic solution concepts often conflict with experimental findings or intuitive reasoning. This fact is possibly due to the requirement that zero probability be assigned to irrational choices in these concepts. Here, we introduce the epistemic notion of common belief in utility proportional beliefs which also assigns positive probability to irrational choices, restricted however by the natural postulate that the probabilities should be proportional to the utilities the respective choices generate. Besides, we propose an algorithmic characterization of our epistemic concept. With regards to experimental findings common belief in utility proportional beliefs fares well in explaining observed behavior.

Keywords: algorithms, epistemic game theory, interactive epistemology, solution concepts, traveler's dilemma, utility proportional beliefs.

1 Introduction

Interactive epistemology, also called epistemic game theory when applied to games, provides a general framework in which epistemic notions such as knowledge and belief can be modeled for situations involving multiple agents. This rather recent discipline has been initiated by Harsanyi (1967-68) as well as Aumann (1976) and first been adopted in the context of games by Aumann (1987), Brandenburger and Dekel (1987) as well as Tan and Werlang (1988). A comprehensive and in-depth introduction to epistemic game theory is provided by Perea (forthcoming). An epistemic approach to game theory analyzes the relation between knowledge, belief, and choice of rational game-playing agents. While classical game theory is based on the two basic primitives – game form and choice – epistemic game theory adds an epistemic framework as a third elementary component such that knowledge and beliefs can be explicitly modelled in games.

* We are grateful to conference participants at the Eleventh Conference of the Society for the Advancement of Economic Theory (SAET2011) as well as to seminar participants at Maastricht University for useful and constructive comments.

Intuitively, an epistemic model of a game can be interpreted as representing the players' reasoning. Indeed, before making a decision in a game, a player reasons about the game and his opponents, given his knowledge and beliefs. Precisely these epistemic mental states on which a player bases his decisions and which characterize his reasoning are described in an epistemically enriched game-theoretic framework.

A central idea in epistemic game theory is common belief in rationality, first explicitly formalized in an epistemic model for games by Tan and Werlang (1988). From an algorithmic perspective it corresponds to rationalizability due to Bernheim (1984) and Pearce (1984). Intuitively, common belief in rationality assumes a player to believe his opponents to choose rationally, to believe his opponents to believe their opponents to choose rationally, etc. However, this basic concept gives counterintuitive as well as experimentally invalidated predictions in some games that have received a lot of attention. Possibly, the requirement that only rational choices are considered and zero probability be assigned to any irrational choice is too strong and does not reflect how real world agents reason.

Here, we introduce the epistemic concept of utility proportional beliefs, according to which a player assigns positive probability also to opponents' irrational choices, while at the same time for every opponent differences in probability must be proportional to difference in utility. In particular, better opponents' choices receive higher probability than inferior choices. Intuitively, probabilities now confer the intrinsic meaning of how good the respective player deems his opponents' choices. The concept of common belief in utility proportional beliefs formalizes the idea that players do not only entertain utility proportional beliefs themselves, but also believe their opponents to do so, believe their opponents to believe their opponents to do so, etc. Philosophically, our concept can be seen as a way of formalizing cautious reasoning, since no choice is excluded from consideration. Rational choice under common belief in utility proportional beliefs fares well with regards to intuition and to explaining experimental findings in games of interest, where classical concepts such as rationalizability perform weakly.

As an illustration consider a simplified version of Basu's (1994) traveler's dilemma. Two persons have traveled with identical items on a plane, however when they arrive their items are damaged and they want to claim compensation by the airline. Both travelers are asked to simultaneously submit a discrete price between 1 and 10. The person with the lower price is then rewarded this value plus a bonus of 2, while the person with the higher price receives the lower price minus a penalty of 2. If the travelers submit the same price, then they both are compensated accordingly. Reasoning in line with common belief in rationality requires a traveler to rationally choose the lowest price 1. Intuitively, the highest price can never be optimal for neither traveler, and iteratively inferring that every respective lower price can then not be optimal neither only leaves the very lowest price as rational choice. However, this conclusion conflicts with experimental findings as well as with intuition, possibly since people typically do not do all iterations or do not assign zero probability to opponents' irrational choices. Indeed, our concept of common belief in utility proportional beliefs leads to the

choice of 6. Intuitively, if all prices receive a substantial positive probability, the very low prices perform quite badly and hence do so too with common belief in utility proportional beliefs.

The basic idea underlying utility proportional beliefs also appears in Rosenthal's (1989) t-solution, where players are required to assign probabilities to their own choices such that the probability differences are proportional to the utility differences using a proportionality factor t. In contrast to our model, Rosenthal uses the same proportionality factor t across all players; assumes that players consciously randomize, i.e. pick probability distributions over their choice sets; and builds in an equilibrium condition implying that players entertain correct beliefs about their opponents' randomized choices.

The intuition that better choices receive higher probabilities also occurs in McKelvey and Palfrey's (1995) quantal response equilibrium, where the utilities are subject to random errors. In contrast to our model, McKelvey and Palfrey do not assume probabilities to be proportional to utilities; require players to hold correct beliefs about the opponents' probabilities; and suppose agents to always choose optimally with respect to their beliefs while their utilities are randomly perturbed.

The scheme of cautious reasoning – that is, no choice is completely discarded from consideration – is also present in Schuhmacher's (1999) and Asheim's (2001) concept of proper rationalizability, which assumes better choices to be infinitely more likely than worse choices. However, in our model every choice receives a substantial, non-infinitesimal positive probability, which is proportional to the utility the respective choice generates.

We proceed as follows. In Section 2, the concept of common belief in utility proportional beliefs is formalized in a type-based epistemic model for games. Also, a convenient way of stating utility proportional beliefs by means of an explicit formula is presented. Rational choice under common belief in utility proportional beliefs is defined as the decision-relevant notion for game-playing agents. Section 3 introduces the algorithm of iterated elimination of utility-disproportional-beliefs, which recursively restricts the players' possible beliefs about the opponents' choices. The algorithm is then shown in Section 4 to yield unique beliefs for every player, which is rather surprising. Section 5 establishes that iterated elimination of utility-disproportional-beliefs provides an algorithmic characterization of common belief in utility proportional beliefs, and can thus be used as a practical tool to compute the beliefs a player can hold when reasoning in line with common belief in utility proportional beliefs. Section 6 illustrates how well our concept fares with regards to intuition as well as experimental findings in some games that have received a lot of attention. Section 7 discusses utility proportional beliefs from a conceptual point of view and compares it to some related literature. Finally, Section 8 offers some concluding remarks and indicates possible directions for future research.

2 Common Belief in Utility Proportional Beliefs

In order to model reasoning in line with utility proportional beliefs, infinite belief hierarchies need to be considered. Here, we restrict attention to static games and follow the type-based approach to epistemic game theory, which represents belief hierarchies as types. More precisely, a set of types is assigned to every player, where each player's type induces a belief on the opponents choices and types. Then, the whole infinite belief hierarchy can be derived from a given type. Note that the notion of type was originally introduced by Harsanyi (1967-68) to model incomplete information, but can actually be more generally used for any interactive uncertainty. Indeed, the context we consider is the uncertainty about choice in finite normal form games.

Notationally, a finite normal form game is represented by the tuple

$$\Gamma = (I, (C_i)_{i \in I}, (U_i)_{i \in I}),$$

where I denotes a finite set of players, C_i denotes player i's finite choice set, and $U_i : \times_{j \in I} C_j \to \mathbb{R}$ denotes player i's utility function.

The notion of an epistemic model constitutes the framework in which various epistemic mental states of players can be described.

Definition 1. *An* epistemic model *of a game Γ is a tuple $\mathcal{M}^\Gamma = ((T_i)_{i \in I}, (b_i)_{i \in I})$, where*

- *T_i is a finite set of types for player $i \in I$,*
- *$b_i : T_i \to \Delta(C_{-i} \times T_{-i})$ assigns to every type $t_i \in T_i$ a probability measure on the set of opponents' choice-type combinations.*

Here, $C_{-i} := \times_{j \in I \setminus \{i\}} C_j$ and $T_{-i} := \times_{j \in I \setminus \{i\}}$ denote the set of opponents' choice and type combinations, respectively. Note that although according to Definition 1 the probability measure $b_i(t_i)$ represents type t_i's belief function on the set of opponents' choice-type pairs, for sake of notational convenience we also use $b_i(t_i)$ to denote any projected belief function for type t_i.[1] It should always be clear from the context which belief function $b_i(t_i)$ refers to.

Besides, in this paper we follow the one-player perspective approach to epistemic game theory advocated by Perea (2007a), (2007b), and (forthcoming). Accordingly, all epistemic concepts including iterated ones are understood and defined as mental states inside the mind of a single person. Indeed, a one-player approach seems natural, since reasoning is formally represented by epistemic concepts and any reasoning process prior to choice takes place entirely within the reasoner's mind.

Some further notions and notation are now introduced. For that purpose consider a game Γ, an epistemic model $\mathcal{M}^\Gamma$ of it, and fix two players $i, j \in I$ such that $i \neq j$. A type $t_i \in T_i$ of i is said to *deem possible* some type $t_j \in T_j$ of

[1] A type's belief function projected on some opponent's type space or projected on the set of opponents' choice combinations are examples for projected belief functions.

his opponent j, if $b_i(t_i)$ assigns positive probability to an opponents' choice-type combination that includes t_j. By $T_j(t_i)$ we then denote the set of types of player j deemed possible by t_i. Furthermore, given a type $t_i \in T_i$ of player i, and given an opponent's type $t_j \in T_j(t_i)$,

$$(b_i(t_i))(c_j \mid t_j) := \frac{(b_i(t_i))(c_j, t_j)}{(b_i(t_i))(t_j)}$$

is type t_i's *conditional belief* that player j chooses c_j given his belief that j is of type t_j. Note that the conditional belief $(b_i(t_i))(c_j \mid t_j)$ is only defined for types of j deemed possible by t_i.

Moreover, a choice combination for player i's opponents is denoted by $c_{-i} \in \times_{j \in I \setminus \{j\}} C_j$. For each of his choices $c_i \in C_i$ type t_i's expected utility given his belief on his opponents' choice combinations is given by

$$u_i(c_i, t_i) = \Sigma_{c_{-i}} (b_i(t_i))(c_{-i}) U_i(c_i, c_{-i}).$$

Besides, let $C := \times_{i \in I} C_i$ be the set of all choice combinations in the game. Then, $\overline{u}_i := max_{c \in C} u_i(c)$ and $\underline{u}_i := min_{c \in C} u_i(c)$ denote the best and worst possible utilities player i can obtain in the game, respectively. Farther, player t_i's average expected utility is denoted by

$$u_i^{average}(t_i) := \frac{1}{\mid C_i \mid} \Sigma_{c_i \in C_i} u_i(c_i, t_i).$$

The idea that a player entertains beliefs on his opponents' choices proportional to the respective utilities these choices yield for the opponents can be formalized within the framework of an epistemic model for normal form games.

Definition 2. *Let $i \in I$ be some player, and $\lambda_i = (\lambda_{ij})_{j \in I \setminus \{i\}} \in \mathbb{R}^{|I \setminus \{i\}|}$ such that $\lambda_{ij} \geq 0$ for all $j \in I \setminus \{i\}$. A type $t_i \in T_i$ of player i expresses* λ_i-utility-proportional-beliefs*, if*

$$(b_i(t_i))(c_j \mid t_j) - (b_i(t_i))(c'_j \mid t_j) = \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j} (u_j(c_j, t_j) - u_j(c'_j, t_j)) \qquad (\star)$$

for all $t_j \in T_j(t_i)$, for all $c_j, c'_j, \in C_j$, for all $j \in I \setminus \{i\}$.

Accordingly, player i's conditional beliefs about two choices of any of his opponents are proportional to the respective utilities the opponent derives from them with proportionality factor $\frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}$ for every opponent $j \in I \setminus \{j\}$.

Note that the variable part of the proportionality factor λ_{ij} in equation $(\star)$ is constrained by the fact that i's belief functions are probability measures. Indeed, λ_i-utility-proportional-beliefs may not exist if any of the λ_{ij}'s is too large. For every opponent $j \in I \setminus \{i\}$, let λ_{ij}^{max} be the highest λ_{ij} such that the system of equations $(\star)$ yields a well-defined probability measure $(b_i(t_i)(\cdot \mid t_j)$ for every type $t_j \in T_j(t_i)$. The tuple $\lambda_i^{max} = (\lambda_{ij}^{max})_{j \in I \setminus \{i\}}$ then contains every opponent's λ_{ij}^{max}.

Intuitively, λ_{ij} measures the sensitivity of the beliefs to differences in utility. If it is too large, some of the beliefs will become negative in order to satisfy equation ($\star$), thus violating the property of being probabilities. Hence, there exists an upper bound for every λ_{ij}, which is captured by λ_{ij}^{max} and furnishes maximally different beliefs as probabilities while at the same time still complying with equation ($\star$) for every opponent's type. Farther, the minimal value λ_{ij} can assume is zero, which then implies the conditional beliefs about the respective opponent's choice to be uniformly distributed. In other words, if $\lambda_{ij} = 0$, then utility differences are not at all reflected in the beliefs, as all choices are being assigned the same probability. Besides, note that in the context of modeling reasoning in line with utility proportional beliefs, choosing $\lambda_{ij} = \lambda_{ij}^{max}$ seems plausible, as the idea of utility proportional beliefs then unfolds its maximal possible effect.

Moreover, λ_i-utility-proportional-beliefs are invariant with respect to affine transformations of any player's utility function. Indeed, suppose $a \in \mathbb{R}$ and $b > 0$ such that $\hat{u}_j(c_j, t_j) = a + bu_j(c_j, t_j)$ for some $j \in I \setminus \{i\}$, for all $c_j \in C_j$, and for all $t_j \in T_j$. Assume that t_i expresses λ_i-utility-proportional-beliefs with respect to $\hat{u}_j$. Then, observe that

$$(b_i(t_i))(c_j \mid t_j) - (b_i(t_i))(c'_j \mid t_j) = \frac{\lambda_{ij}}{\overline{\hat{u}}_j - \underline{\hat{u}}_j}(\hat{u}_j(c_j, t_j) - \hat{u}_j(c'_j, t_j))$$

$$= \frac{\lambda_{ij}}{(a + b\overline{u}_j) - (a + b\underline{u}_j)}((a + bu_j(c_j, t_j)) - (a + bu_j(c'_j, t_j)))$$

$$= \frac{\lambda_{ij}}{b(\overline{u}_j - \underline{u}_j)} b(u_j(c_j, t_j) - u_j(c'_j, t_j))$$

$$= \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j(c'_j, t_j))$$

for all $c_j, c'_j \in C_j$ and for all $t_j \in T_j$. The invariance of utility proportional beliefs with regards to affine transformations of the players' utilities strengthens the concept, since it does not depend on the particular cardinal payoff structure but only on the underlying ordinal preferences.

Utility proportional beliefs as formalized in Definition 2 can be expressed by means of an explicit formula for a given opponent's choice conditional of him being of a given type. This convenient alternative way of stating utility proportional beliefs only relates the conditional belief in a specific opponent's choice to the utilities this choice generates for the respective opponent.

Lemma 1. *Let $i \in I$ be some player, and $\lambda_i = (\lambda_{ij})_{j \in I \setminus \{i\}} \in \mathbb{R}^{|I \setminus \{i\}|}$. A type $t_i \in T_i$ of player i expresses λ_i-utility-proportional-beliefs if and only if*

$$(b_i(t_i))(c_j \mid t_j) = \frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j)),$$

for all $t_j \in T_j(t_i)$, for all $c_j \in C_j$, for all $j \in I \setminus \{i\}$.

Proof. Let $j \in I \setminus \{i\}$ be some opponent of player i, $t_j \in T_j(t_i)$ be some type of j deemed possible by i and $c_j^* \in C_j$ be some choice of j. Note that

$$1 = \Sigma_{c_j \in C_j}(b_i(t_i))(c_j \mid t_j)$$

$$= \Sigma_{c_j \in C_j}((b_i(t_i))(c_j^* \mid t_j) + (b_i(t_i))(c_j \mid t_j) - (b_i(t_i))(c_j^* \mid t_j))$$

$$= \Sigma_{c_j \in C_j}((b_i(t_i))(c_j^* \mid t_j) + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j(c_j^*, t_j)))$$

$$= (\mid C_j \mid (b_i(t_i))(c_j^* \mid t_j) + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}\Sigma_{c_j \in C_j}(u_j(c_j, t_j) - u_j(c_j^*, t_j)))$$

$$= \mid C_j \mid (b_i(t_i))(c_j^* \mid t_j) + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(\mid C_j \mid u_j^{average}(c_j, t_j) - \mid C_j \mid u_j(c_j^*, t_j)),$$

which is equivalent to

$$(b_i(t_i))(c_j^* \mid t_j) = \frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j^*, t_j) - u_j^{average}(t_j)).$$

■

Intuitively, the formula provided by Lemma 1 assigns to every opponents' type the uniform distribution on the respective opponents' choice set plus or minus an adjustment for each choice depending on its goodness relative to the average utility.

In addition to requiring a player to entertain utility proportional beliefs we also assume his beliefs to satisfy *conditional independence*. Intuitively, this condition – due to Brandenburger and Friedenberg (2008) – states that in games with more than two players, the belief that some player i holds about some opponent j's choice must be independent from his belief about some distinct opponent k's choice, if we condition on fixed belief hierarchies of j and k. Conditional independence reflects the natural idea that a player believes his opponents to choose independently from each other, and can formally be defined as follows.

Definition 3. *Let $t_i \in T_i$ be a type for some player $i \in I$. Type t_i holds* conditionally independent beliefs*, if*

$$(b_i(t_i))((c_j, c_k) \mid t_j, t_k) = (b_i(t_i))(c_j \mid t_j) \cdot (b_i(t_i))(c_k \mid t_k)$$

for all $t_j \in T_j(t_i)$, $t_k \in T_k(t_i)$, $c_j \in C_j$, $c_k \in C_k$, and $j, k \in I \setminus \{i\}$ such that $j \neq k$.

Here, $(b_i(t_i))((c_j, c_k) \mid t_j, t_k)$ denotes the conditional probability that t_i assigns to the opponents' choice pair (c_j, c_k), given t_i believes these opponents to be of types t_j and t_k.

Reasoning in line with utility proportional beliefs requires a player not only to entertain utility proportional beliefs himself, but also to believe his opponents to do so, to believe his opponents to believe their opponents to do so, etc. Within an epistemic framework this reasoning assumption can be formally expressed by common belief in utility proportional beliefs.

Definition 4. *Let $i \in I$ be some player, $t_i \in T_i$ be some type of player i, and $\lambda = (\lambda_i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{|I \setminus \{i\}|}$.*

- *Type t_i expresses* 1-fold belief in λ-utility-proportional-beliefs*, if t_i expresses λ_i-utility-proportional-beliefs.*
- *Type t_i expresses k-fold* belief in λ-utility-proportional-beliefs*, if $(b_i(t_i))$ only deems possible types $t_j \in T_j$ for all $j \in I \setminus \{i\}$ such that t_j expresses $k-1$-fold belief in λ-utility-proportional-beliefs, for all $k > 1$.*
- *Type t_i expresses* common belief in λ-utility-proportional-beliefs*, if t_i expresses k-fold belief in λ-utility-proportional-beliefs for all $k \geq 1$.*

Intuitively, a player i expressing common belief in λ-utility-proportional-beliefs holds λ_{ij}-utility-proportional-beliefs on opponent j's choices, he believes opponent j to entertain $\lambda_{jj'}$-utility-proportional-beliefs on opponent j''s choices, etc. In other words, in a belief hierarchy satisfying λ-utility-proportional-beliefs there is no level at which it is not iteratively believed that every opponent holds λ-utility-proportional-beliefs.

Analogously, common belief in conditional independence can be inductively defined. Indeed, the concept we are eventually interested in will be common belief in "λ-utility-proportional-beliefs and conditional independence". From now on common belief in conditional independence is endorsed as an implicit background assumption and hence no longer be explicitly stated.

The choices a player can reasonably make under common belief in utility proportional beliefs are those that are rational under his respectively restricted beliefs on the opponents' choices.

Definition 5. *Let $i \in I$ be some player, and $\lambda = (\lambda_i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{|I \setminus \{i\}|}$. A choice $c_i \in C_i$ of player i is* rational under common belief in λ-utility-proportional-beliefs*, if there exists an epistemic model $\mathcal{M}^{\Gamma}$ and some type $t_i \in T_i$ of player i such that c_i is optimal given $(b_i(t_i))$ and t_i expresses common belief in λ-utility-proportional-beliefs.*

3 Algorithm

An algorithm is introduced that iteratively deletes beliefs and that – as will be shown later in Section 5 – yields precisely those beliefs that are possible under common belief in λ-utility proportional beliefs.

Before we formally define our algorithm some more notation needs to be fixed. Let $P_i^0 := \Delta(C_{-i})$ denote the set of i's beliefs about his opponents' choice combinations. Moreover, for every player i and each of his opponents $j \neq i$ let $p_{ij}^* : P_j^0 \to \Delta(C_j)$ be a function mapping beliefs of player j on his opponents' choice combinations to beliefs on j's choice, defined as

$$(p_{ij}^*(p_j))(c_j) := \frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, p_j) - u_j^{average}(p_j))$$

for all $c_j \in C_j$ and for all $p_j \in P_j^0$. Besides, for every player i let $p_i^* : \times_{j \neq i} P_j^0 \rightarrow P_i^0$ be a function mapping i's opponents' combinations of beliefs to beliefs of i about his opponents, given by

$$(p_i^*((p_j)_{j \neq i}))((c_j)_{j \neq i}) := \Pi_{j \neq i}(p_{ij}^*(p_j))(c_j).$$

The algorithm *iterated elimination of utility-disproportional-beliefs* can now be formally stated.

Definition 6. *Let the sequence of sets $(P_i^k)_{k \geq 0}$ of beliefs on j's choice space be inductively given by*

$$P_i^k := conv(\{p_i^*((p_j)_{j \neq i}) : p_j \in P_j^{k-1} \text{ for all } j \neq i\}).$$

The set of beliefs $P_i^\infty = \bigcap_{k \geq 0} P_i^k$ contains the beliefs that survive iterated elimination of utility-disproportional beliefs.

Intuitively, the function p_i^* transforms beliefs of i's opponents' beliefs on their opponents' choices into beliefs of i on his opponents' choices. The algorithm then iteratively deletes beliefs that cannot be obtained by the functions p_i^*. In other words, beliefs are repeatedly eliminated which are not utility proportional to utilities generated by beliefs from the preceding set of beliefs in the algorithm.

4 Uniqueness of Beliefs

Before it is shown that the iterated utility-disproportional beliefs algorithm converges to singleton sets, some preliminary observations are made. First of all, note that, since λ_{ij}^{max} is the largest possible proportionality factor such that $p_{ij}^*(p_j)$ is a probability distribution for all $p_j \in P_j^0$, it follows that $(p_{ij}^*(p_j))(c_j) \geq 0$ for all $c_j \in C_j$ and for all $p_j \in \Delta(C_i)$ under $\lambda_{ij} = \lambda_{ij}^{max}$. Hence, if $\lambda_{ij} < \lambda_{ij}^{max}$, then $(p_{ij}^*(p_j))(c_j) > 0$ for all $c_j \in C_j$. Moreover, p_{ij}^* is a linear function from P_j^0 to $\Delta(C_j)$, as u_j and $u_j^{average}$ are linear in p_j. Farther, recall that the one norm of a given vector $x = (x_1, x_2, \ldots, x_m)$ is defined as $\|x\|_1 = \Sigma_{i=1}^m \mid x_i \mid$. Notationally, given a player $i \in I$ and some opponents' choice combination $c_{-i} \in C_{-i}$ let $[c_{-i}]$ denote the $\mid C_{-i} \mid$-dimensional vector containing 1 in its c_{-i}-th entry and 0 otherwise, representing a belief on the opponents' choices which puts probability 1 on the choice combination c_{-i}.

The following lemma establishes that p_{ij}^* is a contraction mapping.

Lemma 2. *Let $\lambda_{ij} < \lambda_{ij}^{max}$. There exists $\alpha < 1$ such that $\|p_{ij}^*(p_j) - p_{ij}^*(p_j')\|_1 \leq \alpha \|p_j - p_j'\|_1$ for all $p_j, p_j' \in P_j^0$.*

Proof. As $\lambda_{ij} < \lambda_{ij}^{max}$ it follows that $(p_{ij}^*(p_j))(c_j) > 0$ for all $c_j \in C_j$ and for all $p_j \in P_j^0$. Let $p_{ij}^{min} : P_j^0 \rightarrow \mathbb{R}$ be defined as $p_{ij}^{min}(p_j) := min_{c_j}(p_{ij}^*(p_j))(c_j)$ for all $p_j \in P_j^0$. Note that, since p_{ij}^* is linear, it is also continuous, and thus p_{ij}^{min} as the minimum of finitely many continuous functions is continuous as well. Now, P_j^0

being a non-empty, compact set, the Theorem of Weierstrass ensures that there exists $\epsilon > 0$ such that $(p_{ij}^{min}(p_j)) \geq \epsilon$ for all $p_j \in P_j^0$ and hence $(p_{ij}^*(p_j))(c_j) \geq \epsilon$ for all $c_j \in C_j$ and for all $p_j \in P_j^0$, too.

We now show that

$$\|p_{ij}^*(p_j) - p_{ij}^*(p_j')\|_1 \leq (1-\epsilon)\|p_j - p_j'\|_1.$$

Observe that due to the linearity of p_{ij}^*, it holds that

$$(p_{ij}^*(p_j))(c_j) = \Sigma_{c_{-j} \in C_{-j}} p_j(c_{-j})(p_{ij}^*([c_{-j}]))(c_j)$$

for all $c_j \in C_j$ and for all $p_j \in P_j^0$. It then follows that

$$\|p_{ij}^*(p_j) - p_{ij}^*(p_j')\|_1 = \Sigma_{c_j \in C_j} \mid (p_{ij}^*(p_j))(c_j) - (p_{ij}^*(p_j'))(c_j) \mid$$

$$= \Sigma_{c_j \in C_j} \mid \Sigma_{c_{-j} \in C_{-j}} p_j(c_{-j})(p_{ij}^*([c_{-j}]))(c_j) - \Sigma_{c_{-j} \in C_{-j}} p_j'(c_{-j})(p_{ij}^*([c_{-j}]))(c_j) \mid$$

$$= \Sigma_{c_j \in C_j} \mid \Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j) \mid .$$

For some fixed $c_j \in C_j$ consider

$$\Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j)$$

and note that either

$$\Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j) \geq 0$$

or

$$\Sigma_{c_{-j} \in C_{-j}} (p_j'(c_{-j}) - p_j(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j) \geq 0.$$

Without loss of generality assume that

$$\Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j) \geq 0.$$

Recalling that $(p_{ij}^*([c_{-j}]))(c_j) \geq \epsilon$ for all $c_j \in C_j$ and for all $c_{-j} \in C_{-j}$, it thus follows that,

$$\mid \Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j) \mid$$

$$= \Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j}))(p_{ij}^*([c_{-j}]))(c_j)$$

$$= \Sigma_{c_{-j} \in C_{-j} : p_j(c_{-j}) - p_j'(c_{-j}) \geq 0} \mid p_j(c_{-j}) - p_j'(c_{-j}) \mid (p_{ij}^*([c_{-j}]))(c_j)$$

$$-\Sigma_{c_{-j} : p_j(c_{-j}) - p_j'(c_{-j}) < 0} \mid p_j'(c_{-j}) - p_j(c_{-j}) \mid (p_{ij}^*([c_{-j}]))(c_j)$$

$$\leq \Sigma_{c_{-j} \in C_{-j} : p_j(c_{-j}) - p_j'(c_{-j}) \geq 0} \mid p_j(c_{-j}) - p_j'(c_{-j}) \mid (p_{ij}^*([c_{-j}]))(c_j)$$

$$-\epsilon \Sigma_{c_{-j} : p_j(c_{-j}) - p_j'(c_{-j}) < 0} \mid p_j'(c_{-j}) - p_j(c_{-j}) \mid .$$

Note that

$$\Sigma_{c_{-j} \in C_{-j}} (p_j(c_{-j}) - p_j'(c_{-j})) = 0$$

due to

$$\Sigma_{c_{-j}\in C_{-j}} p_j(c_{-j}) = 1 = \Sigma_{c_{-j}\in C_{-j}} p'_j(c_{-j}).$$

It hence follows that

$$\Sigma_{c_{-j}:p_j(c_{-j})<p'_j(c_{-j})} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid = \Sigma_{c_{-j}:p_j(c_{-j})>p'_j(c_{-j})} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid .$$

Since it is also the case that

$$\mid p_j(c_{-j}) - p'_j(c_{-j}) \mid ((p^*_{ij}([c_{-j}]))(c_j) - \epsilon) \geq 0$$

for all $c_{-j} \in C_{-j}$ and for all $c_j \in C_j$,

$$\Sigma_{c_{-j}\in C_{-j}:p_j(c_{-j})-p'_j(c_{-j})\geq 0} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid (p^*_{ij}([c_{-j}]))(c_j)$$

$$-\epsilon\Sigma_{c_{-j}:p_j(c_{-j})-p'_j(c_{-j})<0} \mid p'_j(c_{-j}) - p_j(c_{-j}) \mid$$

$$= \Sigma_{c_{-j}\in C_{-j}:p_j(c_{-j})-p'_j(c_{-j})\geq 0} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid ((p^*_{ij}([c_{-j}]))(c_j) - \epsilon)$$

$$\leq \Sigma_{c_{-j}\in C_{-j}} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid ((p^*_{ij}([c_{-j}]))(c_j) - \epsilon)$$

therefore obtains. Consequently,

$$\|p^*_{ij}(p_j) - p^*_{ij}(p'_j)\|_1 = \Sigma_{c_j\in C_j} \mid \Sigma_{c_{-j}\in C_{-j}} (p_j(c_{-j}) - p'_j(c_{-j}))(p^*_{ij}([c_{-j}]))(c_j) \mid$$

$$\leq \Sigma_{c_j\in C_j}\Sigma_{c_{-j}\in C_{-j}} \mid p_j(c_{-j}) - p'_j(c_{-j}) \mid ((p^*_{ij}([c_{-j}]))(c_j) - \epsilon)$$

$$= \Sigma_{c_{-j}\in C_{-j}}(\mid (p_j(c_{-j}) - p'_j(c_{-j}))\Sigma_{c_j\in C_j}((p^*_{ij}([c_{-j}]))(c_j) - \epsilon) \mid).$$

As

$$\Sigma_{c_j\in C_j}(p^*_{ij}([c_{-j}]))(c_j) = 1$$

for all $c_{-j} \in C_{-j}$, it follows that

$$\Sigma_{c_{-j}\in C_{-j}}(\mid (p_j(c_{-j}) - p'_j(c_{-j}))\Sigma_{c_j\in C_j}((p^*_{ij}([c_{-j}]))(c_j) - \epsilon) \mid)$$

$$= \Sigma_{c_{-j}\in C_{-j}}(\mid (p_j(c_{-j}) - p'_j(c_{-j}))(1 - \mid C_j \mid \epsilon) \mid)$$

$$= (1 - \mid C_j \mid \epsilon)\|p_j - p'_j\|_1 \leq (1 - \epsilon)\|p_j - p'_j\|_1.$$

Hence,

$$\|p^*_{ij}(p_j) - p^*_{ij}(p'_j)\|_1 \leq (1 - \epsilon)\|p_j - p'_j\|_1$$

holds and since $\epsilon > 0$ the desired result obtains. ■

Note that for the result of Lemma 2 it does not matter which particular norm is employed. In fact, we used $\|\cdot\|_1$ as it facilitates the proof.

For the following theorem some additional notation is needed. Recall that P^k_i is the set of beliefs for player i about the opponents' choices that survive k rounds of the algorithm. For every two distinct players $i \neq j$ let P^k_{ij} denote the projection of P^k_i on $\Delta(C_j)$.

Theorem 1. *Let* $M := max_{i \in I}\{max_{p_i, \hat{p}_i \in P_i^0}\{\|p_i - \hat{p}_i\|_1\}\}$. *There exists* $\alpha < 1$ *such that* $\|p'_{ij} - p''_{ij}\|_1 \leq \alpha^k M$ *for all* $p'_{ij}, p''_{ij} \in P_i^k$, *for all* $k \geq 0$, *and for all* $i \neq j$.

Proof. First of all, note that by Lemma 2, for every two distinct players i, j there exists $\alpha_{ij} < 1$ such that $\|p^*_{ij}(p_j) - p^*_{ij}(p'_j)\|_1 \leq \alpha_{ij}\|p_j - p'_j\|_1$ for all $p_j, p'_j \in P_j^0$. Take α to be the maximum of all these α_{ij}.

The proof proceeds by induction on k. Observe that $\|p'_i - p''_i\|_1 \leq M$ for all $p'_i, p''_i \in P_i^0$. Now, let $p'_{ij}, p''_{ij} \in P_i^{k+1}$ and note that $p'_{ij} = p^*_{ij}(p'_j)$ and $p''_{ij} = p^*_{ij}(p''_j)$ for some $p'_j, p''_j \in P_j^k$. With Lemma 2 and the induction hypothesis, it then follows that

$$\|p'_{ij} - p''_{ij}\|_1 = \|p^*_{ij}(p'_j) - p^*_{ij}(p''_j)\|_1$$

$$\leq \alpha\|p'_j - p''_j\|_1 \leq \alpha\alpha^k M = \alpha^{k+1} M.$$

■

This preceeding Theorem can be used to show that the beliefs the algorithm yields are unique.

Theorem 2. *Let* $\lambda_{ij} < \lambda_{ij}^{max}$ *for all distinct players* i *and* j. *Then,* $\mid P_i^\infty \mid = 1$ *for all players* i.

Proof. It follows directly from Theorem 1 that P_i^∞ and P_j^∞ contain exactly one belief vector, respectively. ■

The uniqueness of beliefs – which intuitively follows from recursively applying a contraction mapping – constitutes a highly convenient property of the algorithm.

Equipped with Theorem 3 the algorithm can be restated in simplified way, if $\lambda_{ij} < \lambda_{ij}^{max}$ for all distinct players i and j. In that case we can do without taking the convex hull at each step in the algorithm.

Corollary 1. *Let* $\lambda_{ij} < \lambda_{ij}^{max}$ *for all distinct players* i *and* j. *Let the sequence of sets* $(\hat{P}_i^k)_{k \geq 0}$ *of beliefs be inductively given by*

$$\hat{P}_i^k := (\{p_i^*((p_j)_{j \neq i}) : p_j \in \hat{P}_j^{k-1} \text{ for all } j \neq i\}).$$

Then, $\hat{P}_i^\infty = P_i^\infty$ *for all players* i.

Proof. Since by construction $\hat{P}_i^k$ is a subset of P_i^k for all $k > 0$, it follows that $\hat{P}_i^\infty$ is a subset of P_i^∞. As P_i^∞ contains exactly one belief vector, $\hat{P}_i^\infty$ must be equal to P_i^∞. ■

Note that Corollary 1 provides a highly convenient formulation of our algorithm for application purposes. Besides, in the case of two players the equivalence of both algorithms ensues directly, as the sets P_i^k are already convex for all $k > 0$.

5 Algorithmic Characterization of Common Belief in Utility Proportional Beliefs

It is now established that the algorithm yields precisely those beliefs that a player can entertain under common belief in λ-utility proportional beliefs.

Theorem 3. *Let $\lambda = (\lambda_i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{|I \setminus \{i\}|}$ such that $\lambda_{ij} < \lambda_{ij}^{max}$ for all two distinct players i and j. A belief $p_i \in \Delta(C_{-i})$ can be held by a type $t_i \in T_i$ that expresses common belief in λ-utility-proportional beliefs in some epistemic model $\mathcal{M}^{\Gamma}$ of Γ if and only if p_i survives iterated elimination of utility-disproportional beliefs.*

Proof. For the *only if* direction of the theorem, we prove by induction on k that a belief that can be held by a type that expresses up to k-fold belief in λ-utility-proportional beliefs survives k rounds of iterated elimination of utility-disproportional beliefs. It then follows, that a type expressing common belief in λ-utility-proportional-beliefs holds a belief which survives iterated elimination of utility-disproportional beliefs.

First of all, let $k = 1$ and consider $t_i \in T_i$ that expresses 1-fold belief in λ-utility-proportional beliefs. Then,

$$(b_i(t_i))(c_j, t_j) = (\frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j)))(b_i(t_i))(t_j)$$

for all $c_j \in C_j$, $t_j \in T_j$ and for all $j \in I \setminus \{i\}$. It follows that,

$$(b_i(t_i))(c_j) = \Sigma_{t_j \in T_j(t_i)}(b_i(t_i))(t_j)(\frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j)))$$

for all $c_j \in C_j$. Written as a vector,

$$(b_i(t_i))(c_j))_{c_j \in C_j} = \Sigma_{t_j \in T_j(t_i)} b_i(t_i))(t_j)(\frac{1}{\mid C_j \mid}(1, \cdots, 1)$$

$$+ \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j))_{c_j \in C_j})$$

obtains. Note that by definition of the algorithm,

$$\frac{1}{\mid C_j \mid}(1, \cdots, 1) + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j))_{c_j \in C_j} \in P_i^1$$

holds. Since $(b_i(t_i))(c_j))_{c_j \in C_j}$ thus is a convex combination of elements in the convex set P_i^1, it follows that $(b_i(t_i))(c_j))_{c_j \in C_j} \in P_i^1$.

Now let $k \geq 1$ and consider $t_i \in T_i$ that expresses k-fold belief in λ-utility-proportional beliefs. Then,

$$(b_i(t_i))(c_j, t_j) = (\frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j)))(b_i(t_i))(t_j)$$

for all $c_j \in C_j$ and for all $t_j \in T_j$. Therefore,

$$(b_i(t_i))(c_j) = \Sigma_{t_j \in T_j(t_i) \subseteq B_j^{k-1}} (b_i(t_i))(t_j)(\frac{1}{\mid C_j \mid} + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j)))$$

for all $c_j \in C_j$, where B_j^{k-1} denotes the set of j's types that express $k-1$-fold belief in λ-utility-proportional-beliefs. Written as a vector,

$$(b_i(t_i))(c_j))_{c_j \in C_j} = \Sigma_{t_j \in T_j(t_i) \subseteq B_j^{k-1}} b_i(t_i))(t_j)(\frac{1}{\mid C_j \mid}(1, \cdots, 1)$$

$$+ \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j))_{c_j \in C_j})$$

obtains. Since every $t_j \in T_j(t_i)$ is in B_j^{k-1} and hence by the induction hypothesis $(b_j(t_j)) \in P_j^{k-1}$, it follows that

$$\frac{1}{\mid C_j \mid}(1, \cdots, 1) + \frac{\lambda_{ij}}{\overline{u}_j - \underline{u}_j}(u_j(c_j, t_j) - u_j^{average}(t_j))_{c_j \in C_j}$$

$$= p_i^*(b_j(t_j)) \in P_i^k.$$

As $(b_i(t_i))(c_j))_{c_j \in C_j}$ thus is a convex combination of elements in the convex set P_i^k, $(b_i(t_i))(c_j))_{c_j \in C_j} \in P_i^k$ holds as well. By induction on k the *only if* direction of the theorem follows.

For the *if* direction of the theorem, let $p_i \in P_i^\infty$ and $p_j \in P_j^\infty$, which are unique by Theorem 2, respectively. Consider the epistemic model

$$\mathcal{M}^\Gamma = ((T_i, T_j), (b_i, b_j))$$

of Γ, where $T_i = \{t_i\}$, $T_j = \{t_j\}$, $b_i(t_i)$ projected on T_j puts probability 1 on t_j, while projected on C_j equals p_i. Furthermore, $b_j(t_j)$ projected on T_i puts probability 1 on t_i, while projected on c_{-j} equals $p_j(p_i)$]. Note that $(b_i(t_i))(c_j \mid t_j) = (b_i(t_i))(c_j)$ for all $c_j \in C_j$ and $(b_j(t_j))(c_{-j} \mid t_i) = (b_j(t_j))(c_{-j})$ for all $c_{-j} \in c_{-j}$, since $\mid T_i \mid = 1 = \mid T_j \mid$. Using the fact that

$$p_i = p_i^*(p_j) \text{ and } p_j = p_j^*(p_i),$$

it then follows, by construction of b_i and b_j, that both t_i as well as t_j express λ-utility-proportional-beliefs. Moreover, t_i and t_j express common belief in λ-utility-proportional-beliefs too, as $T_j(t_i) = \{t_j\}$ and $T_i(t_j) = \{t_i\}$. ■

According to the preceding theorem the algorithm thus provides a convenient way to compute the beliefs a player can hold when reasoning in line with common belief in utility proportional beliefs.

Farther, note that the proof of the *if* direction of Theorem 3 establishes that common belief in utility proportional beliefs is always possible in every game.

Corollary 2. *Let $\lambda_i = (\lambda_{ij})_{j \in I \setminus \{i\}} \in \mathbb{R}^{|I \setminus \{i\}|}$. There exists an epistemic model $\mathcal{M}^\Gamma$ of Γ, and a type $t_i \in T_i$ for every every player $i \in I$ such that t_i expresses common belief in λ-utility-proportional-beliefs.*

It is thus guaranteed that common belief in utility proportional beliefs is a logically sound concept, which can be adopted to describe players' reasoning in any game.

Linking the algorithmic characterization of common belief in utility proportional beliefs with Theorem 2 establishes the uniqueness of the beliefs that can be held under common belief in utility proportional beliefs.

Theorem 4. *Let $\lambda_{ij} < \lambda_{ij}^{max}$ and $\lambda_{ji} < \lambda_{ji}^{max}$. For both players, there exists a unique belief about the opponent's choice that he can hold under common belief in λ-utility-proportional-beliefs.*

Proof. By Theorem 2 the algorithm yields unique beliefs for both players, and Theorem 3 ensures that these are precisely the beliefs that they can hold under common belief in λ-utility-proportional-beliefs. ■

Consequently, a player reasoning in line with common belief in utility proportional beliefs can only entertain a unique belief about his opponents' choices. The suitability of the concept to be used for descriptions in games is therefore very high.

6 Illustration

Due to its property of always providing unique beliefs for the players about their respective opponents' choices, the algorithm is easy and conveniently implementable. Indeed, without any difficulties we wrote a small program, which computes this unique belief vector given a 2-player game in normal form. We now illustrate in some well-known games that have received a lot of attention how well common belief in utility proportional beliefs fares with respect to intuition as well as to experimental findings – in contrast to classical concepts which run into problems when applied to these games. In each example we use λ_{ij} slightly smaller than λ_{ij}^{max} such that the differences in utilities have the largest possible effect on the players' beliefs, while still guaranteeing these beliefs to be unique. In fact, from these unique beliefs it is possible to directly read off the rational choices under common belief in λ-utility-proportional-beliefs, since these choices must receive the highest probability under those beliefs.

Example 1. Consider again the traveler's dilemma which has already been introduced in Section 1. Consider three variants of the game, according to which the players can choose between 10, 30, and 100 prices. Reasoning in line with common belief in rationality requires the travelers to opt for the minimum price of 1 in each of the three variations. However, it neither seems plausible to exclude any irrational choice completely from consideration nor do experiments confirm such results. For instance, in an experiment with members of the game theory

society by Becker et al (2005), where prices between 2 and 100 could be chosen, most persons opted for a high price of at least 90. In fact, contrary to common belief in rationality, our concept yields the much more natural choices of 6, 26, and 96, respectively. Besides, note that common belief in utility proportional beliefs is actually sensitive to the cardinality of the choice sets. Indeed, it seems intuitive that when there are few prices to choose from rather lower prices will be opted for, and when there are many prices available then the ones picked will be higher. ♣

Example 2. Figure 2 depicts an asymmetric matching pennies game that is taken from Goeree and Holt (2001).

		Column Player	
		$left$	$right$
Row Player	top	$320, 40$	$40, 80$
	$bottom$	$40, 80$	$80, 40$

Fig. 1. Asymmetric matching pennies

In the unique Nash equilibrium of the game, *Row Player* chooses $(\frac{1}{2}, \frac{1}{2})$ and *Column Player* chooses $(\frac{7}{8}, \frac{1}{8})$. Intuitively, it seems reasonable for *Row Player* to opt for *top* due to the very high possible payoff of 320, while *Column Player* might tend to pick *right* anticipating *Row Player*'s temptation for *top*. Indeed, in experiments by Goeree and Holt (2001) approximately 95 % of the row players choose *top*, while approximately 85 % of the column players opt for *right*. Here, close to the experimental findings our concept of common belief in utility proportional beliefs yield choices *top* and *right* for *Row Player* and *Column Player*, respectively. ♣.

Example 3. Suppose the normal form in Figure 3 which models a coordination game with a secure outside option that is taken from Goeree and Holt (2001).

		Column Player		
		$left$	$middle$	$right$
Row Player	top	$90, 90$	$0, 0$	$0, 40$
	$bottom$	$0, 0$	$180, 180$	$0, 40$

Fig. 2. A coordination game with a secure outside option

The game contains multiple Nash equilibria, among which there is the focal high-payoff one $(bottom, middle)$, while *Column Player* has access to a secure outside option guaranteeing him a payoff of 40. In experiments by Goeree and Holt (2001), approximately 95 % of the row players choose *bottom*, while approximately 95 % of the column players pick *middle*. Close to the results from

the laboratory, common belief in utility proportional beliefs yields *bottom* and *middle*. ♣

Example 4. The Kreps Game due to Kreps (1995) is represented in Figure 4.

		Column Player			
		left	*middle*	*non-Nash*	*right*
Row Player	*top*	$20, 5$	$0, 4$	$1, 3$	$2, -10^4$
	bottom	$0, -10^4$	$1, -10^3$	$3, 3$	$5, 10$

Fig. 3. Kreps game

The game exhibits three Nash equilibria, two pure and one mixed, and in none of them *Column Player* chooses *Non-Nash* with positive proabability. However, *Non-Nash* appears to be reasonable, as all other options only yield a slightly higher payoff, but might lead to considerable losses. If anticipating this reasoning *Row player* would optimally choose *bottom*. Indeed, in informal experiments by Kreps the row players pick *bottom*, while the column players choose *Non-Nash* in the majority of cases. Also in this game common belief in utility proportional beliefs performs intuitively by generating *top* for *Row Player* and *non-Nash* for *Column Player*, respectively. Indeed, *top* seems reasonable for the *Row Player* as long as he assigns a substantial probability to the *Column Player* choosing *left*, which is what our concept does. ♣

7 Discussion

Utility Proportional Beliefs. The concept of utility proportional beliefs seems quite a natural and basic way of reasoning in the context of games. Indeed, it does appear plausible to assign non-zero probability to opponents' irrational choices due to causes such as the complexity of the interactive situation, uncertainty about the opponents' utilities and choice rules, possibility of mistakes, caution etc. However, at the same time it is intuitive that the opponents' relative utilities are reflected in a player's beliefs about their choice and to thus assign probabilities proportional to the respective utilities.

Moreover, utility proportional beliefs furnishes probabilities with intrinsic meaning in the sense of measuring how good a player deems some choice for the respective opponent, and thus also provides an account of *how* agents form their beliefs. In contrast, basic classical concepts like common belief in rationality treat every choice that receives positive probability as equally plausible.

Besides, utility proportional beliefs does not only appear reasonable from intuitive as well as theoretical perspectives, but also fares well with regards to experimental findings, as indicated in Section 6. In this context, also note that experimental findings can often not be explained by the basic concept of common belief in rationality, which implies that any irrational choice always

receives zero probability. However, if it is assumed that players follow a common line of reasoning, then positive probability *must* be assigned to irrational choices, which is precisely what utility proportional beliefs does.

t-Solutions. Rosenthal's (1989) class of t-solutions for two player games formalizes the idea that players do not exclusively play best responses. Intuitively, given a fixed parameter $t \in \mathbb{R}$, a pair of randomized choices constitutes a t-solution, if each of them satisfies the property that if positive probability is assigned to some pure choice, then the difference in probability with any other pure choice of the same player equals t times the difference in the respective utilities given the opponent's randomized choice.[2] In other words, players assign probabilities to their choices such that the probability differences are proportional to the utility difference multiplied by the proportionality factor t.

In contrast to our concept of utility proportional beliefs, Rosenthal's t-solutions employs a proportionality factor which is the same across all players. It seems more desirable to permit different agents to entertain distinct proportionality factors, in order to represent heterogenous states of mind, and to thus provide a more realistic account of reasoning. Also, t-solutions are not invariant to affine translations of the utilities, which is a serious drawback not arising in our model. Moreover, the players' probability distributions which are restricted by a utility proportionality condition are distinct objects in Rosenthal's and our models. While in the former randomized choices, i.e. conscious randomizations of the players are considered, beliefs on the opponents' choices are used in the latter. Since assuming probabilities to be objects of choice constitutes a problematic assumption for at least most game-theoretic contexts, probabilities interpreted as players' beliefs seems more plausible and realistic. Besides, by keeping the opponents' choices fixed, an equilibrium condition is built into Rosenthal's t-solution concept. However, from an epistemic point of view fixing the opponents' choices seems highly unreasonable, as it means that the reasoner already knows what his opponents will do in the game. Note that in our model we admit players to be erroneous about their opponents' choices as well as beliefs, which again is closer to real life, where people are frequently not correct about their fellow men's choices in interactive situations. Furthermore, if the randomized choices of a player are interpreted as his opponents' beliefs about his choice, and if the same proportionality factor is applied to all players, then with the uniqueness of the beliefs result of Theorem 2, our concept becomes equivalent to Rosenthal's t-solution. Note that the uniqueness of the beliefs imply Rosenthal's equilibrium condition. However, we do not impose the equilibrium condition, but it follows as a result from our model.

Quantal Response Equilibrium. McKelvey and Palfrey (1995) introduce the concept of quantal response equilibrium as a statistical version of equilibrium, where each player chooses deterministically, however his utility for each of his choices

[2] Given a game Γ and a player $i \in I$, a *randomized choice* for i is a probability distribution $\sigma_i \in \Delta(C_i)$ on i's choice space.

is subject to random error. Given a rational decision rule players are assumed to apply, the random error induces a probability distribution over the players' observed choices. In their model these probabilities satisfy the intuitive property that better choices are more likely to be chosen than worse choices.

In contrast to our concept of utility proportional beliefs, McKelvey and Palfrey do not require the probability of a given choice to be *proportional* to the expected utility it generates. Yet, in terms of reasoning it appears natural that a player assigns utility proportional probabilities to his opponents' choices – as in our model – when deliberating about what his opponents might choose. Moreover, the probabilities in quantal response equilibrium are not invariant to affine translations of the utilities. This serious drawback is avoided in our model. Besides, from an epistemic point of view McKelvey and Palfrey's equilibrium condition implicitly assumes that players' know their opponents' random error induced probabilities. This seems rather implausible, as players can never have direct access to opponents' minds. Farther, in McKelvey and Palfrey's model agents are assumed to always choose best responses with respect to their beliefs but not with respect to their utilities , which are randomly perturbed. However, in our model the utilities are kept fixed, but we allow players to assign positive probability to opponents' suboptimal choices.

Proper Rationalizability. The concept of proper rationalizability, introduced by Schumacher (1999) as well as Asheim (2001), and algorithmically characterized by Perea (2011), formalizes cautious reasoning in games. Intuitively, a choice is properly rationalizable for a player, if he is cautious, i.e. does not exclude any opponent's choice from consideration; respects his opponents' preferences, i.e. if he believes an opponent to prefer some choice c to c', then he deems c infinitely more likely than c'; as well as expresses common belief in the event that his opponents' are cautious and respect their opponents' preferences. A standard tool to model infinitely-more-likely relations are lexicographic beliefs.[3] Loosely speaking, a reasoner is then said to be cautious, if for every opponent each of his choices occur in the support of the probability distribution of some lexicographic level of the reasoner's lexicographic belief. Hence, all opponents' choices receive positive probability somewhere in a cautious lexicographic belief.

In the sense of modeling cautious reasoning that considers all choices including irrational ones, proper rationalizability and utility proportional beliefs are similar concepts. However, on the one hand, utility proportional beliefs can be viewed as a milder version than proper rationalizability, since the former assigns substantial, infinitesimal positive probability to any choice including non-optimal ones, while the latter assigns infinitesimal probabilities to non-optimal choices. On the other hand, the two concepts can be viewed as opposite ways of cautious reasoning, since utility proportional beliefs reflects the utility differences of choices, while proper rationalizability treats all non-optimal choices as infinitely less likely than optimal choices. Moreoever, on the purely formal level both ways

[3] Given some set W a *lexicographic belief* is a finite sequence $\rho = (\rho^1, \rho^2, \ldots, \rho^K)$ of probability distributions such that $\rho^k \in \Delta(W)$ for all $k \in \{1, 2, \ldots, K\}$.

of cautious reasoning are distinct, as utility proportional beliefs employs standard beliefs, whereas proper rationalizability models lexicographically minded agents.

8 Conclusion

Utility proportional beliefs provides a basic and natural way of of reasoning in games. The underlying intuitions that irrational choices should not be completely neglected, and beliefs ought to reflect how good a player deems his opponents' choices, seem plausible. The surprising property that the iterated elimination of utility-disproportional-beliefs algorithm yields unique beliefs strengthens the suitability of common belief in utility proportional beliefs to be used for descriptions in games. Moreover, in various games of interest our concept matches well intuition and experimental findings.

The idea of utility proportional beliefs opens up a new direction of research. Naturally, the concept can be extended to dynamic games. Besides, the effects of allowing uncertainty about the opponents' λ's can be studied. Moreover, applications of our epistemic concept to well-known games or economic problems such as auctions might be highly interesting.

References

Asheim, G. B. (2001): Proper Rationalizability in Lexicographic Beliefs. *International Journal of Game Theory*, 30, 453–478.

Aumann, R. J. (1976): Agreeing to Disagree. *Annals of Statistics* 4, 1236–1239.

Aumann, R. J. (1987): Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica* 55, 1–18.

Basu, K. (1994): The Traveler's Dilemma: Paradoxes of Rationality in Game Theory. *American Economic Review* 84, 391–395.

Becker, T., Carter, M., and Naeve, J. (2005): Experts Playing the Traveler's Dilemma. Discussion Paper 252/2005, Universitt Hohenheim.

Bernheim, B. D. (1984): Rationalizable Strategic Behavior. *Econometrica* 52, 1007–1028.

Brandenburger, A. and Dekel, E. (1987): Rationalizability and Correlated Equilibria. *Econometrica* 55, 1391–1402.

Brandenburger, A. and Friedenberg, A. (2008): Intrinsic Correlation in Games. *Journal of Economic Theory* 141, 28–67.

Goeree, J. K. and Holt, C. A. (2001): Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *American Economic Review* 91, 1402–1422.

Harsanyi, J. C. (1967-68): Games of Incomplete Information played by "Bayesian Players". Part I, II, III. *Management Science* 14, 159–182, 320–334, 486–502.

Kreps, D. M. (1995): Nash Equilibrium. In *The New Palgrave: Game Theory*, Norton, New York, pp. 176-177.

McKelvey, R. D. and Palfrey, T. R. (1995): Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10, 6–38.

Pearce, D. (1984): Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica* 52, 1029–1050.

Perea, A. (2007a): A One-Person Doxastic Characterization of Nash Strategies. *Synthese*, 158, 1251–1271.

Perea, A. (2007b): Epistemic Conditions for Backward Induction: An Overview. In *Interactive Logic Proceedings of the 7th Augustus de Morgan Workshop, London. Texts in Logic and Games 1.* Amsterdam University Press, pp. 159–193.

Perea, A. (2011): An Algorithm for Proper Rationalizability. *Games and Economic Behavior*, 72, 510–525.

Perea, A. (forthcoming): *Epistemic Game Theory: Reasoning and Choice.* Cambridge University Press.

Rosenthal, R. W. (1989): A Bounded-Rationality Approach to the Study of Noncooperative Games. *International Journal of Game Theory*, 18, 273–292.

Schuhmacher, F. (1999): Proper Rationalizability and Backward Induction. *International Journal of Game Theory*, 28, 599–615.

Tan, T. C. C. and Werlang, S. R. C. (1988): The Bayesian Foundation of Solution Concepts of Games. *Journal of Economic Theory* 45, 370–391.