Towards Formal Foundations for Game Theory

Julian Parsert(✉) and Cezary Kaliszyk

Department of Computer Science, University of Innsbruck, Innsbruck, Austria
{julian.parsert,cezary.kaliszyk}@uibk.ac.at

Abstract. Utility functions form an essential part of game theory and economics. In order to guarantee the existence of these utility functions sufficient properties are assumed in an axiomatic manner. In this paper we discuss these axioms and the von-Neumann-Morgenstern Utility Theorem, which names precise assumptions under which expected utility functions exist. We formalize these results in Isabelle/HOL. The formalization includes formal definitions of the underlying concepts including continuity and independence of preferences. We make the dependencies more precise and highlight some consequences for a formalization of game theory.

1 Introduction

Utility theory seeks to describe how humans evaluate and compare alternatives or outcomes using mathematical tools. This theory forms the basis of game theory and therefore several fields in economics. Hence, we believe that formalizations in either of those areas require a solid base in utility theory.

In their pioneering work "Theory of Games and Economic Behavior" von Neumann and Morgenstern axiomatically describe, how actors evaluate *uncertain outcomes* [22]. They developed the theory of expected utility, which describes a scheme based on the expected value of outcomes. Utility functions allow the use of many mathematical tools for optimization etc. Hence, much effort is put into precisely specifying properties which guarantee the existence of such functions. To this end, von Neumann and Morgenstern dedicate the first chapters of [22] to specifying the assumptions necessary (and sufficient) for preference relations to admit expected utility representation. This is now known as the von-Neumann-Morgenstern Utility Theorem. These assumptions are introduced as *axioms* upon which the entire book is based. Kahneman and others criticized [20] the theory of expected utility and developed alternatives [7]. Moreover, impossibility results were proven [19]. Nevertheless, it still remains the standard theory in game theory [14] and the most common tool in economic reasoning [9].

Our goal is to provide a solid foundation of utility theory upon which further work in both economics and game theory can be conducted. We do so by introducing formal definitions in Isabelle/HOL and deriving results that not

J. Avigad and A. Mahboubi (Eds.): ITP 2018, LNCS 10895, pp. 495–503, 2018.
https://doi.org/10.1007/978-3-319-94821-8_29

only support the intuition of expected utility, but also help automated theorem provers in proving subsequent results. With that we prove the von-Neumann-Morgenstern Expected Utility Theorem.

Related Work. Arrow's impossibility theorem has been formalized by Wiedijk [25] and Nipkow [13]. Gammie has formalized some results in social choice theory, as well as stable matching [4,5]. Kuhn's theorem has been formalized by Vestergaard [21] and generalized by Le Roux [17]. The same author later worked on a formalization of Nash equilibria for two player games [18]. Recently, Martin-Dorel and Soloviev formalized boolean games with non-deterministic aspects. In addition, algorithmic game theory results have been formalized in Coq [1].

The concepts we discuss are also relevant for the formalization of economic concepts. Related work includes the verification of financial systems [16] and binomial pricing models [3]. As part of the ForMaRE project [10] VCG-Auctions [8] have been formalized. In microeconomics we discussed a formalization of two economic models and the First Welfare Theorem [15].

To our knowledge the only work that uses expected utility theory is that of Eberl [2]. The focus there is not the underlying utility theory, but rather its use in social decision schemes. Since our focus is the this underlying theory and in particular the von Neumann-Morgenstern Utility Theorem, we found that there is only little overlap.

2 Isabelle/HOL, Probability, and Notations

Isabelle/HOL [24] is an *Interactive Theorem Prover* based on higher-order logic. Due to space limitations, we refer the reader to the Isar reference manual [23] for Isabelle's foundations and notations. We introduce a few reoccurring notions of HOL-Probability, but we refer to [6] for a more detailed explanation.

It is common to denote the composition of probability mass functions (pmfs) p and q with a probability α as follows $\alpha\ p + (1-\alpha)\ q$. This notation corresponds to the following Isabelle definition:

$$\textbf{definition}\ \texttt{mix\_pmf} :: \mathit{real} \Rightarrow {'a}\ \mathit{pmf} \Rightarrow {'a}\ \mathit{pmf} \Rightarrow {'a}\ \mathit{pmf}\ \textbf{where}$$
$$\texttt{mix\_pmf}\ a\ p\ q\ =\ (\mathit{bernoulli} - \mathit{pmf}\ a) \gg\!= (\lambda b.\ \textbf{if}\ b\ \textbf{then}\ p\ \textbf{else}\ q)$$

In particular, we compose a Bernoulli distribution that returns either *True* or *False* with probability a, with a function that returns p if the random variable is *True* or q otherwise. We use Isabelle's standard definition for the support of a pmf, `set_pmf`, while `return_pmf` applied to x returns a pmf yielding x with the probability 1.

A preference relation is a transitive and reflexive binary relation (i.e. a preorder). The notations $x \succeq y$, $x \succeq [R]\ y$, and $R(x, y)$ are equivalent and denote a preference relation where x is weakly preferred to y. Despite its potential ambiguity, we will be using the first alternative if the specific relation can be inferred

from context. Similarly, the symbols $x \succ y$ and $x \succ[R]\ y$ denote the strict preference relation where $x \succ y$ iff $x \succeq y \wedge \neg y \succeq x$, whereas $x \approx y$ and $x \approx[R]\ y$ denote the indifference relation where $x \approx y$ iff $x \succeq y \wedge y \succeq x$.

We will use the terms "pmf" and "lottery" interchangeably. In economic and game theoretic literature the latter is more common, while the former is used in probability theory and Isabelle/HOL.

3 Preference Relations and Their Properties

We present and discuss important definitions which we will use in subsequent sections.

First we briefly introduce rational preferences and Utility functions. However, since both have been thoroughly discussed and formalized in the authors' previous work [15] we will not go into detail or mention results involving these.

Definition 1 (Rational Preferences). *A binary relation R over a carrier set C is called a rational preference relation, if R is a total preorder on C. Hence R is total, transitive, and reflexive.*

We refer to [15] or the sources for a more detailed account of the Definitions 1 and 2 as well as derived results.

Definition 2 (Utility function). *A function $u : C \mapsto \mathbb{R}$ is said to represent a rational preference relation R over C, if*

$$\forall x\ y \in C.\ \ x \succeq[R]\ y \iff u(x) \geq u(y).$$

The function u is called utility function.

Based on these two definitions we continue with the new additions. Firstly, we consider continuous preferences. Definition 3 is sometimes also called the *Archimedean axiom*.

Definition 3 (Continuous Preferences). *A binary relation R over a carrier set C, is said to be continuous if, $\forall\ p\ q\ r \in C$,*

$$p \succeq[R]\ q \wedge q \succeq[R]\ r \longrightarrow \exists \alpha \in \left[0 \ldots 1\right].(\texttt{mix\_pmf}\ \alpha\ p\ r) \approx[R]\ q.$$

Intuitively this means that if $p \succ q$, then lotteries that are *close* to p are also preferred to q. An alternative interpretation would be, that if preferences are continuous, there are no outcomes that are so bad (not preferred with respect to R) that no probability is small enough to "redeem" them by composing with a better alternative.

Next, we define independence of preferences. Informally, we want independence to entail that the (preference) relation between two elements p and q only depends on the parts where p and q differ.

Definition 4 (Independence of Preferences). *A binary relation R over a carrier set C, is independent if,* $\forall p\ q\ x \in C.\ \forall \alpha \in (0 \ldots 1]$,

$$p \succeq[R]\ q \longleftrightarrow (\texttt{mix\_pmf}\ \alpha\ p\ x) \succeq[R]\ (\texttt{mix\_pmf}\ \alpha\ q\ x).$$

Independence implies that the relation between $\alpha\ p + (1-\alpha)\ x$ and $\alpha\ q + (1-\alpha)\ x$ only depends on the relation of p and q rather than their combination with x.

Even though utility functions have been defined, the special case of *expected* utility functions has not been discussed. We will do so now.

Definition 5 (Expected Utility Form[1]). *Given a set P of probability mass functions over a set of outcomes $\mathcal{O}$ and a preference relation R over P, a utility function $\mathcal{U} : P \mapsto \mathbb{R}$ representing R has expected utility form, if there exists a utility function u[2] $: \mathcal{O} \mapsto \mathbb{R}$ such that for all $p \in P$,*

$$\mathcal{U}(p) := \sum_{x \in \mathcal{O}} p(x) * u(x).$$

Notice that Definition 5 introduces two kinds of utility functions, the expected utility function $\mathcal{U}$ and the Bernoulli utility function u. The function $\mathcal{U}$ assigns a utility value to lotteries/pmfs that *range over* outcomes, while u assigns a utility value to outcomes themselves. The utility of a lottery p which equals $\mathcal{U}(p)$ is then defined to be the expected value of the utility function u with the lottery p.

4 The Setup

In this section we introduce notations that we use and discuss further concepts and assumptions.

First, we assume the set of outcomes $\mathcal{O}$ to be a non-empty finite set[3]. Next, we define the carrier set $\mathcal{P}$ to be the set of all probability mass functions (pmf) over the finite set of outcomes $\mathcal{O}$, $\mathcal{P} := \{l \mid \mathit{support}\ l \subseteq \mathcal{O}\}$. This set can be visualized using a probability simplex. Figure 1 shows such a simplex with three outcomes. Note, that if $|\mathcal{O}| > 1$ then the set P is uncountable. Now, we can define *degenerate lotteries* to be all lotteries that yield one outcome with the probability 1. In Fig. 1 these are simply the corner points (i.e., the points $\mathcal{O}_{1-3}$). A rational preference relation over $\mathcal{P}$ is denoted with $\mathcal{R}$. Since the final result requires $\mathcal{R}$ to be continuous and independent (cf. Definitions 3 and 4) most literature assumes these from the get go. We found that not all assumptions were necessary for the results. Therefore, in the formalization we chose to introduce assumptions only when necessary. Nevertheless, for the sake of readability we assume $\mathcal{R}$ to be rational (1), continuous (3), and independent (4) in the subsequent sections. For more detail on the necessity of assumptions we refer to the formalization.

With this setup, we can state the theorem we are aiming for, the von-Neumann-Morgenstern Utility Theorem (Theorem 1).

[1] This form is also known as the von-Neumann-Morgenstern utility function.
[2] This function is sometimes referred to as Bernoulli utility function.
[3] The discussed theorem also holds for infinite sets [9]. However, this has not been formalized.

Theorem 1 (von-Neumann-Morgenstern Utility Theorem). *The preference relation $\mathcal{R}$ over the carrier set $\mathcal{P}$ can be represented by a utility function of expected utility form (Definition 5) if and only if $\mathcal{R}$ is rational (1), continuous (3), and satisfies independence (4). More formally, $\mathcal{R}$ satisfies (1), (3), (4), if and only if, $\exists u : \mathcal{O} \mapsto \mathbb{R}$ such that $\forall p\ q \in \mathcal{P}$,*

$$p \succeq q \iff \sum_{x \in \mathcal{O}} p(x) * u(x) \geq \sum_{x \in \mathcal{O}} q(x) * u(x).$$

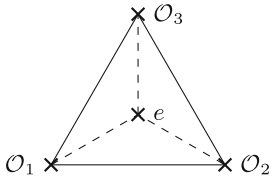

Fig. 1. This is the probability simplex for the case where $|\mathcal{O}| = 3$. The set $\{l \mid \textit{support}\ l \subseteq \mathcal{O}\}$ is exactly the set of all points on this simplex. The point e is the pmf with the probability $\frac{1}{3}$ for all three outcomes $(\frac{1}{3}\mathcal{O}_1 + \frac{1}{3}\mathcal{O}_2 + \frac{1}{3}\mathcal{O}_3)$.

5 The Proof Outline

We will present the key insights and ideas leading to a proof of Theorem 1. All the definitions and proofs can be found in the formalization. Since we use the setup introduced in the previous section all assumptions and notations carry over. In particular $\succeq$ will denote the previously introduced relation $\mathcal{R}$.

Theorem 1 is proved by showing two implications. Both directions can be found in the formalization. However, we will discuss the more difficult direction. That is, a preference relation satisfying (1), (3), and (4) admits expected utility representation.

The set of degenerate lotteries is finite, trivially there exists at least one most preferred element (with respect to $\mathcal{R}$). Moreover, we can prove Lemma 1.

Lemma 1. *Every best*[4] *degenerate lottery B_{deg} is at least as good as any other lottery in $\mathcal{P}$.*

$$\forall y \in \mathcal{P}.\ B_{deg} \succeq y$$

The same can be shown for the worst (least preferred) elements. Thus proving that there exists at least one best $\mathcal{B}$ and one worst $\mathcal{W}$ element in $\mathcal{P}$ such that

$$\forall x \in \mathcal{P}.\ \mathcal{B} \succeq x \wedge x \succeq \mathcal{W}. \tag{1}$$

[4] We will use the "best" and "worst" to denote most and least preferred with respect to $\mathcal{R}$.

If $\mathcal{B} \approx \mathcal{W}$ any constant function would represent the preference relation $\mathcal{R}$, thus proving Theorem 1 for this special case. Hence, we will assume $\mathcal{B} \succ \mathcal{W}$.

From the assumption of continuity and Property 1, we know that $\forall p \in \mathcal{P}$,

$$\exists \alpha.\ \alpha\ \mathcal{B} + (1 - \alpha)\ \mathcal{W} \approx p.$$

Moreover, we can show that such an α is unique. If it was not, we could create two distinct lotteries $p = \alpha\ \mathcal{B} + (1 - \alpha)\ \mathcal{W}$ and $q = \beta\ \mathcal{B} + (1 - \beta)\ \mathcal{W}$ with $\alpha > \beta$ and $p \approx q$. However, since $\mathcal{B} \succ \mathcal{W}$ and p has a higher chance of the best outcome than q, we deduce $p \succ q$, a contradiction. This shows that for all lotteries $p \in \mathcal{P}$, there exists a unique *calibration probability* α, such that, $\alpha\mathcal{B} + (1 - \alpha)\mathcal{W} \approx p$.

The key idea is to define a function that assigns the unique calibration probability to every lottery in $\mathcal{P}$. This is realised with the utility function `util`. Given a pmf p its unique calibration α is obtained (using the indefinite choice operator `SOME`) and returned.

definition `util` :: *'a pmf* ⇒ *real* **where**
  `util` p = (*SOME* α. $\alpha \in \{0 \ldots 1\} \wedge p \approx[\mathcal{R}]$ `mix_pmf` $\alpha\ \mathcal{B}\ \mathcal{W}$)

The next lemma shows that `util` indeed is a utility function as per Definition 2.

Lemma 2. *For all p and q in $\mathcal{P}$,*

$$p \succeq q \iff \texttt{util}(p) \geq \texttt{util}(q)$$

Lemma 2 is already an important result. However, since we are not only interested in general utility functions, but utility functions that adhere to expected utility form (Definition 5), we also need to prove the following Lemma.

Lemma 3. `util` *is linear. That is, for all p, q in $\mathcal{P}$,*

$$\texttt{util}(\alpha\ p + (1 - \alpha)\ q) = \alpha\ \texttt{util}(p) + (1 - \alpha)\ \texttt{util}(q)$$

Proof Outline. First, we generate two lotteries that have the same preference as p and q using `util`, $\mathcal{B}$, and $\mathcal{W}$. After substituting these generated lotteries in the left hand side of the equation, we can distribute α, rearrange the terms and apply the definition of `util` to derive the right hand side. For a detailed account of this lemma, we refer to the formalization. □

One of the most prominent modern books on game theory [11] defines von-Neumann-Morgenstern utility functions simply as linear functions which `util` indeed is (Lemma 3). Since linearity is the defining property of expected utility functions Lemma 4 can be proven. Note, that `util` has the wrong type *'α pmf ⇒ real*. Therefore, we simply define the Bernoulli utility function u with the following lambda abstraction (λx. `util`(`return_pmf` x)) of type *'α ⇒ real*.

Lemma 4. *Given a $p \in \mathcal{P}$*

$$\mathcal{U}(p) = \sum_{x \in \mathcal{O}} p(x) * u(x)$$

This shows the existence of an expected utility function assuming (1), (3), and (4), thus proving one direction of Theorem 1.

6 Conclusions

As mentioned in Sect. 1 multiple prominent books including [11,22], introduce the theory of expected utility as a set of axioms upon which their work is based. Thus, a formalization of utility theory is crucial for further development in game theory and economics. The presented formalization amounts to almost 2400 lines of code including over 120 lemmas. These can be used for future work such as Nash's theorem [12] on the existence of mixed strategy equilibria.

Acknowledgments. We thank Manuel Eberl for his help with Isabelle's HOL-Probability. This work is supported by the European Research Council (ERC) grant no 714034 *SMART* and the Austrian Science Fund (FWF) project P26201.

References

1. Bagnall, A., Merten, S., Stewart, G.: A library for algorithmic game theory in Ssreflect/Coq. J. Formalized Reasoning **10**(1), 67–95 (2017). https://jfr.unibo.it/article/view/7235
2. Eberl, M.: Randomised social choice theory. Archive of Formal Proofs, May 2016. http://isa-afp.org/entries/Randomised_Social_Choice.shtml. Formal proof development
3. Echenim, M., Peltier, N.: The binomial pricing model in finance: a formalization in Isabelle. In: de Moura, L. (ed.) CADE 2017. LNCS (LNAI), vol. 10395, pp. 546–562. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63046-5_33
4. Gammie, P.: Some classical results in social choice theory. Archive of Formal Proofs, November 2008. http://isa-afp.org/entries/SenSocialChoice.html. Formal proof development
5. Gammie, P.: Stable matching. Archive of Formal Proofs, October 2016. http://isa-afp.org/entries/Stable_Matching.html. Formal proof development
6. Hölzl, J.: Construction and stochastic applications of measure spaces in higher-order logic. Ph.D. thesis, Technical University Munich (2013). http://nbn-resolving.de/urn:nbn:de:bvb:91-diss-20130219-1116512-0-6
7. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979). http://www.jstor.org/stable/1914185
8. Kerber, M., Lange, C., Rowat, C., Windsteiger, W.: Developing an auction theory toolbox. AISB, pp. 1–4 (2013)
9. Kreps, D.: Notes on the Theory of Choice. Underground Classics in Economics. Avalon Publishing (1988). https://books.google.at/books?id=9D0Oljs5GrQC
10. Lange, C., Rowat, C., Kerber, M.: The ForMaRE project – formal mathematical reasoning in economics. In: Carette, J., Aspinall, D., Lange, C., Sojka, P., Windsteiger, W. (eds.) CICM 2013. LNCS (LNAI), vol. 7961, pp. 330–334. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39320-4_23
11. Maschler, M., Solan, E., Zamir, S.: Game Theory. Cambridge University Press, New York (2013)
12. Nash, J.F.: Equilibrium points in n-person games. Proc. Nat. Acad. Sci. U.S.A. **36**, 48–49 (1950)
13. Nipkow, T.: Arrow and Gibbard-Satterthwaite. Archive of Formal Proofs (2008). https://www.isa-afp.org/entries/ArrowImpossibilityGS.shtml

14. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V.V.: Algorithmic Game Theory. Cambridge University Press, New York (2007)
15. Parsert, J., Kaliszyk, C.: Formal microeconomic foundations and the first welfare theorem. In: Proceedings of the 7th ACM SIGPLAN International Conference on Certified Programs and Proofs, CPP 2018, pp. 91–101. ACM (2018). https://doi.org/10.1145/3167100
16. Passmore, G.O., Ignatovich, D.: Formal verification of financial algorithms. In: de Moura, L. (ed.) CADE 2017. LNCS (LNAI), vol. 10395, pp. 26–41. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63046-5_3
17. Roux, S.: Acyclic preferences and existence of sequential nash equilibria: a formal and constructive equivalence. In: Berghofer, S., Nipkow, T., Urban, C., Wenzel, M. (eds.) TPHOLs 2009. LNCS, vol. 5674, pp. 293–309. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03359-9_21
18. Roux, S.L., Martin-Dorel, É., Smaus, J.: An existence theorem of Nash equilibrium in Coq and Isabelle. In: Bouyer, P., Orlandini, A., Pietro, P.S. (eds.) Proceedings Eighth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2017, EPTCS, vol. 256, Roma, Italy, 20–22 September 2017, pp. 46–60 (2017). https://doi.org/10.4204/EPTCS.256.4
19. Roux, S.L., Pauly, A.: Extending finite-memory determinacy to multi-player games. Inf. Comput. (2018). http://www.sciencedirect.com/science/article/pii/S0890540118300270
20. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. Science **185**(4157), 1124–1131 (1974). http://science.sciencemag.org/content/185/4157/1124
21. Vestergaard, R.: A constructive approach to sequential nash equilibria. Inf. Process. Lett. **97**(2), 46–51 (2006). https://doi.org/10.1016/j.ipl.2005.09.010
22. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press (1947). https://books.google.at/books?id=AUDPAAAAMAAJ
23. Wenzel, M.: The Isabelle/Isar Reference Manual (2017)
24. Wenzel, M., Paulson, L.C., Nipkow, T.: The Isabelle framework. In: Mohamed, O.A., Muñoz, C., Tahar, S. (eds.) TPHOLs 2008. LNCS, vol. 5170, pp. 33–38. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71067-7_7
25. Wiedijk, F.: Formalizing Arrow's theorem. Sadhana **34**(1), 193–220 (2009). https://doi.org/10.1007/s12046-009-0005-1