Consequentialism in Dynamic Games*

Andrés Perea†

EPICENTER Working Paper No. 30 (2024)

Abstract

In philosophy and decision theory, *consequentialism* reflects the assumption that an act is evaluated solely on the basis of the consequences it may induce, and nothing else. In this paper we study the idea of consequentialism in dynamic games by considering two versions: A commonly used utility-based version stating that the player's preferences are governed by a utility function on consequences, and a preference-based version which faithfully translates the original idea of consequentialism to restrictions on the player's preferences. Utility-based consequentialism always implies preference-based consequentialism, but the other direction is not necessarily true, as is shown by means of a counterexample. It turns out that utility-based consequentialism is equivalent to the assumption that the induced preference intensities on consequences are additive, whereas preference-based consequentialism only requires this property for every pair of strategies in isolation. We finally show that if the dynamic game either (i) has two strategies for the player we consider, or (ii) has observed past choices, or (iii) involves only two players and has perfect recall, then the two notions of consequentialism are equivalent in the absence of weakly dominated strategies.

JEL classification: C72, D81

Keywords: Consequentialism, utility-based consequentialism, preference-based consequentialism, dynamic games, conditional preference relation, additive preference intensities on consequences.

*I would like to thank two referees at the LOFT 2024 conferences, and the seminar audiences at Maastricht University and Universidad Autónoma de Madrid for their valuable feedback.

†*E-mail:* a.perea@maastrichtuniversity.nl *Affiliation:* Maastricht University, Department of Quantitative Economics and EpiCenter, The Netherlands.

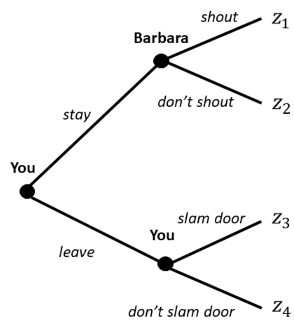

Figure 1: Illustration of consequentialism

1 Introduction

In philosophy and decision theory, *consequentialism* reflects the assumption that a person evaluates an act solely based on the possible consequences that this particular act may induce, and nothing more. For a detailed account the reader may consult the overviews by Sinnott-Armstrong (2023) and Machina (1989, Section 4), and the references therein.

In the game theoretic literature the notion of consequentialism has rarely been discussed explicitly. However, the dynamic games we traditionally use implicitly assume a strong version of consequentialism, by writing down utilities at the terminal histories, or consequences, and assuming that the player's preferences are governed by such utilities. We refer to this assumption as *utility-based* consequentialism. It is also assumed in many well-known decision theoretic models such as von Neumann and Morgenstern (1944), Savage (1954) and Anscombe and Aumann (1963). Indeed, in these models the proposed axioms guarantee that the decision maker's preferences can be represented by a utility function on consequences, supplemented in Savage (1954) and Anscombe and Aumann (1963) by a subjective belief on states.

We argue, however, that there are natural scenarios of dynamic games where utility-based consequentialism is violated. As an illustration, consider the situation where you have a discussion with your friend Barbara. After a calm start the discussion has entered a stage where you must decide between *staying, leaving the room while slamming the door,* and *leaving the room calmly.* If you stay, Barbara has the option to either start shouting at you or to teach you a lesson without raising her voice. This leads to the dynamic game form depicted in Figure 1.

Suppose you are determined to leave, but you still have to decide whether to slam the door or not. If you are a consequentialist, your preference between these two options should not depend on whether you believe that Barbara would (counterfactually) start to shout or not if you were to stay, as it does not affect the consequences (z_3 or z_4) of the two options mentioned above. However, one could still imagine a scenario where you would be prone to slam the door if you believe that Barbara would counterfactually start shouting at you if you were to stay, whereas you would prefer to leave calmly if you believe that Barbara would not start shouting at you in that situation. As such, it makes perfect sense not be a consequentialist in this game.

The question we wish to investigate in this paper is to what extent the notion of utility-based consequentialism faithfully represents the original idea of consequentialism as described at the beginning of the introduction. Or is this notion too strong in some scenarios? To address this question we use the decision theoretic framework by Gilboa and Schmeidler (2003) and Perea (2023), which requires the decision maker to hold a *conditional preference relation* assigning to every probabilistic belief over the states a preference relation over his acts. The reason we use this framework is that it naturally fits the analysis of games. Indeed, if we apply it to dynamic games, then a player is supposed to hold a preference relation over his own strategies for every possible probabilistic belief about the opponents' strategies. This naturally reflects the game theoretic element that the ranking of your own strategies crucially depends on what you believe that others will do.

Within this decision theoretic setting we formulate a *preference-based* version of consequentialism which states that the ranking of two strategies under a given belief should only depend on the probability distributions over consequences induced by these two strategies under the belief, and nothing more. It is therefore a faithful translation of the original idea of consequentialism to the setting of dynamic games.

It turns out that utility-based consequentialism always implies preference-based consequentialism, but the other direction may not be true. We offer an example of a three-player game where past choices are imperfectly observed such that a particular player satisfies preference-based, but not utility-based, consequentialism.

The difference between the two notions in this example is that utility-based consequentialism induces *additive* preference intensities on consequences for this player, whereas preference-based consequentialism does not. By additive preference intensities on consequences we mean that for every three consequences x, y and z, the sum of the intensity by which you prefer x to y and the intensity by which you prefer y to z equals the intensity by which you prefer x to z. In fact, we show in Theorem 4.1 that for *every* dynamic game form, utility-based consequentialism can be characterized by the condition that the conditional preference relation at hand induces preference intensities on consequences that are additive. In turn, Theorem 4.2 states that preference-based consequentialism is equivalent to the weaker requirement that the property above holds for every *pair* of strategies in isolation, but not necessarily for all strategies together.

In Theorem 5.1 we identify conditions under which the two notions of consequentialism are equivalent. More precisely, it is shown that if the dynamic game form either (i) has only two strategies for the player under consideration, or (ii) has observed past choices, or (iii) has only two players and satisfies perfect recall, then the two notions of consequentialism are equivalent, provided there is an expected utility representation for the conditional preference relation and there are no weakly dominated strategies. For such scenarios, the condition of additive induced preference intensities on consequences is thus implied by preference-based consequentialism alone. These are precisely the situations where writing down utilities at the terminal histories faithfully reflects the original idea of consequentialism.

The outline of this paper is as follows: In Section 2 we introduce our model of a dynamic game and the decision theoretic framework as described above. In Section 3 we lay out the two definitions of consequentialism. In Section 4 we provide an example where the two notions of consequentialism are not equivalent, prove that utility-based consequentialism is equivalent to requiring that the induced preference intensities on consequences are additive, and that preference-based consequentialism is equivalent to requiring that the induced preference intensities on consequences are additive for every pair of strategies in isolation. In Section 5 we identify a set of sufficient conditions under which the two notions of consequentialism are equivalent. In Section 6 we provide some concluding remarks. The appendix contains the proofs of the three theorems, together with some definitions from graph theory, some preparatory results, and a utility transformation procedure, which are needed for the proofs.

2 Model

In this section we start by laying out our model of a dynamic game form, followed by the definition of a strategy and that of a conditional preference relation for a distinguished player.

2.1 Dynamic Game Forms

In this paper we consider finite dynamic games that allow for simultaneous moves and imperfect information. Formally, a *dynamic game form* is a tuple $D = (I, P, I^a, (A_i, H_i)_{i \in I}, Z)$, where

(a) I is the finite set of *players*;

(b) P is the finite set of *past action profiles*, or *histories;*

(c) the mapping I^a assigns to every history $p \in P$ the (possibly empty) set of *active players* $I^a(p) \subseteq I$ who must choose after history p. If $I^a(p)$ contains more than one player, there are simultaneous moves after p. If $I^a(p)$ is empty, the game terminates after p. By P_i we denote the set of histories $p \in P$ with $i \in I^a(p)$;

(d) for every player i, the mapping A_i assigns to every history $p \in P_i$ the finite set of *actions* $A_i(p)$ from which player i can choose after history p. The objects P, I^a and $(A_i)_{i \in I}$ must be such that the empty history $\emptyset$ is in P, representing the beginning of the game, and the non-empty histories in P are precisely those objects $(p, (a_i)_{i \in I^a(p)})$ where p is a history in P and $a_i \in A_i(p)$ for every $i \in I^a(p)$;

(e) for every player i there is a partition H_i of the set of histories P_i where i is active. Every partition element $h_i \in H_i$ is called an *information set* for player i. In case h_i contains more than one history, the interpretation is that player i does not know at h_i which history in h_i has been reached. The objects A_i and H_i must be such that for every information set $h_i \in H_i$ and every two histories p, p' in h_i, we have that $A_i(p) = A_i(p')$. We can thus write $A_i(h_i)$ for the unique set of available actions at h_i. Moreover, it must be that $A_i(h_i) \cap A_i(h_i') = \emptyset$ for every two distinct information sets $h_i, h_i' \in H_i$;

(f) $Z \subseteq P$ is the collection of histories p where the set of active players $I^a(p)$ is empty. Such histories are called *terminal* histories, or *consequences.*

This definition follows Osborne and Rubinstein (1994), with the difference that we do not specify utilities at the consequences. This is why we call it a dynamic game *form* and not a dynamic game.

Based on this model we can derive the following definitions: We say that a history p *precedes* a history p' (or p' *follows* p) if p' results by adding some action profiles after p. Let $H := \cup_{i \in I} H_i$ be the collection of all information sets for all players. For every two information sets $h, h' \in H$, we say that h *precedes* h'

(or h' *follows* h) if there is a history $p \in h$ and a history $p' \in h'$ such that p precedes p'. Two information sets h, h' are *simultaneous* if there is some history p which belongs to both h and h'. We say that h *weakly precedes* h' (or h' *weakly follows* h) if either h precedes h', or h, h' are simultaneous.

The dynamic game form satisfies *perfect recall* (Kuhn (1953)) if every player always remembers which actions he chose in the past, and which information he had about the opponents' past actions. Formally, for every player i, every information set $h \in H_i$, and every two histories $p, p' \in H_i$, the sequence of player i actions in p and p' must be the same (and consequently, the collection of player i information sets that p and p' cross must be the same).

The dynamic game form has *observed past choices,* also known as *observable actions,* if every player always observes all choices that have been made in the past. Formally, for every player i, every information set $h_i \in H_i$ consists of a single history.

2.2 Strategies

A strategy for player i assigns an available action to every information set at which player i is active, and that is not excluded by earlier actions in the strategy. Formally, let $\tilde{s}_i$ be a mapping that assigns to *every* information set $h_i \in H_i$ some action $\tilde{s}_i(h) \in A_i(h)$. We call $\tilde{s}_i$ a *complete strategy.* Then, a history $p \in P$ is *excluded* by $\tilde{s}_i$ if there is some information set $h_i \in H_i$, with some history $p' \in h_i$ preceding p, such that $\tilde{s}_i(h_i)$ is different from the unique player i action at p' leading to p. An information set $h \in H$ is excluded by $\tilde{s}_i$ if all histories in h are excluded by $\tilde{s}_i$. The *strategy* induced by $\tilde{s}_i$ is the restriction of $\tilde{s}_i$ to those information sets in H_i that are not excluded by $\tilde{s}_i$. A mapping $s_i : \tilde{H}_i \to \cup_{h \in \tilde{H}_i} A_i(h)$, where $\tilde{H}_i \subseteq H_i$, is a *strategy* for player i if it is the strategy induced by a complete strategy.[1] By S_i we denote the set of strategies for player i, and by $S_{-i} := \times_{j \neq i} S_j$ the set of strategy combinations for i's opponents.

Consider a strategy profile $s = (s_i)_{i \in I}$ in $\times_{i \in I} S_i$. Then, s induces a unique consequence $z(s)$. We say that the strategy profile s *reaches* a history p if p precedes $z(s)$. Similarly, the strategy profile s is said to reach an information set h if s reaches a history in h.

For a given information set $h \in H$ and player i we define the sets

$$\begin{aligned} S(h) &:= \{s \in \times_{i \in I} S_i \mid s \text{ reaches } h\}, \\ S_i(h) &:= \{s_i \in S_i \mid \text{there is some } s_{-i} \in S_{-i} \text{ such that } (s_i, s_{-i}) \in S(h)\}, \text{ and} \\ S_{-i}(h) &:= \{s_{-i} \in S_{-i} \mid \text{there is some } s_i \in S_i \text{ such that } (s_i, s_{-i}) \in S(h)\}. \end{aligned}$$

Intuively, $S_i(h)$ is the set of strategies for player i that allow for information set h to be reached, whereas $S_{-i}(h)$ is the set of opponents' strategy combinations that allow for h to be reached.

It is well-known that under perfect recall we have, for every player i and every information set $h_i \in H_i$, that $S(h_i) = S_i(h_i) \times S_{-i}(h_i)$, and that under observed past choices it holds that $S(h) = \times_{i \in I} S_i(h)$ for every information set h.

[1] What we call a "strategy" is sometimes called a "plan of action" in the literature (Rubinstein (1991)), and what we call a "complete strategy" is often called a "strategy".

For a given strategy $s_i \in S_i$, we denote by $H_i(s_i) := \{h_i \in H_i \mid s_i \in S_i(h_i)\}$ the collection of information sets for player i that the strategy s_i allows to be reached. Similarly, for a given strategy combination $s_{-i} \in S_{-i}$ and a player j, we denote by $H_j(s_{-i}) := \{h_j \in H_j \mid s_{-i} \in S_{-i}(h_j)\}$ the collection of information sets for player j that the strategy combination s_{-i} allows to be reached.

2.3 Conditional Preference Relations

Consider a dynamic game form D and a distinguished player i. Then, the *acts,* or objects of choice, for player i are his strategies in S_i, whereas the *states,* or the events about which he is uncertain, are the opponents' strategy combinations in S_{-i}. Following Gilboa and Schmeidler (2003) and Perea (2023), player i holds for every probabilistic belief about the states a preference relation over his acts. In the definition below we denote by $\Delta(S_{-i})$ the set of probability distributions over S_{-i}.

Definition 2.1 (Conditional preference relation) *For a given dynamic game form D, a **conditional preference relation** for player i is a mapping $\succsim_i$ which assigns to every belief $\beta_i \in \Delta(S_{-i})$ over the opponents' strategy combinations a complete, transitive and reflexive preference relation $\succsim_{i,\beta_i}$ over the strategies in S_i.*

This concept reflects the crucial game theoretic element that player i's ranking over his strategies depends on the belief he holds about the opponents' strategies.

For a given conditional preference relation $\succsim_i$ and two strategies s_i, t_i, we say that s_i *weakly dominates* t_i under $\succsim_i$ if $s_i \succsim_{i,\beta_i} t_i$ for every belief β_i, and $s_i \succ_{i,\beta_i} t_i$ for some belief β_i.

3 Two Notions of Consequentialism

In this section we introduce the preference-based and utility-based versions of consequentialism.

3.1 Preference-Based Consequentialism

We call a conditional preference relation *preference-based consequentialist* if for the ranking of two strategies under a given belief, the player only pays attention to the probability distributions over *consequences* induced by these two strategies under that particular belief. To define it formally we need the following piece of notation: For a given strategy s_i and belief $\beta_i \in \Delta(S_{-i})$, the induced *probability distribution* $\mathbb{P}_{(s_i,\beta_i)}$ *over consequences* is given by

$$\mathbb{P}_{(s_i,\beta_i)}(z) := \sum_{s_{-i} \in S_{-i} : z(s_i,s_{-i}) = z} \beta_i(s_{-i})$$

for all consequences $z \in Z$.

Definition 3.1 (Preference-based consequentialism) *A conditional preference relation $\succsim_i$ is **preference-based consequentialist** if for every four strategies s_i, s_i', t_i, t_i' (not necessarily pairwise different) and every two beliefs β_i and β_i' (not necessarily different) with*

$$\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i')} \text{ and } \mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i',\beta_i')}$$

it holds that

$$s_i \succsim_{i,\beta_i} t_i \text{ if and only if } s_i' \succsim_{i,\beta_i'} t_i'.$$

This definition is similar to the notion of *probabilistic sophistication* in Machina and Schmeidler (1992) and Grant (1995), which states that within the Savage framework, the decision maker holds a unique probabilistic belief over states, and compares two acts only on the basis of their induced probability distributions over consequences.

As an illustration, let us go back to the dynamic game form in Figure 1. Suppose you are player 1 and Barbara is player 2. In the definition above choose the belief β_1 for you that assigns probability 1 to Barbara *shouting,* the belief β_1' that assigns probability 1 to Barbara *not shouting,* the strategy $s_1 = s_1' =$ (*leave, slam door*) and the strategy $t_1 = t_1' =$ (*leave, don't slam door*). Then, $\mathbb{P}_{(s_1,\beta_1)} = \mathbb{P}_{(s_1',\beta_1')}$ and $\mathbb{P}_{(t_1,\beta_1)} = \mathbb{P}_{(t_1',\beta_1')}$. Hence, if you are a preference-based consequentialist, then $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1 \succsim_{1,\beta_1'} t_1$. That is, your preference between (*leave, slam door*) and (*leave, don't slam door*) should not depend on the belief you have about Barbara's counterfactual attitude if you were to stay. This matches precisely the original idea of consequentialism.

3.2 Utility-Based Consequentialism

Following Gilboa and Schmeidler (2003) and Perea (2023), we say that a conditional preference relation has an *expected utility representation* if there is a utility function, assigning to every act-state pair some utility, such that for every belief the decision maker prefers act a to act b precisely when the first act induces a higher expected utility than the second.

Definition 3.2 (Expected utility representation) *Consider a conditional preference relation $\succsim_i$ and a utility function $u_i : S_i \times S_{-i} \to \mathbf{R}$. Then, u_i is an **expected utility representation** for $\succsim_i$ if for every belief $\beta_i \in \Delta(S_{-i})$, and every two strategies s_i, t_i, we have that $s_i \succsim_{i,\beta_i} t_i$ if and only if*

$$\sum_{s_{-i} \in S_{-i}} \beta_i(s_{-i}) \cdot u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \beta_i(s_{-i}) \cdot u_i(t_i, s_{-i}).$$

If this expected utility representation assigns the same utility to any two strategy combinations that induce the same consequence, then we say that the conditional preference relation is *utility-based consequentialist.*

Definition 3.3 (Utility-based consequentialism) *A conditional preference relation $\succsim_i$ is **utility-based consequentialist** if it has an expected utility representation u_i such that for every two strategies s_i, t_i and*

	shout	*don't shout*
stay	2	0
leave, slam door	3	2
leave, don't slam door	5	4

Table 1: Expected utility representation in game of Figure 1

every two opponents' strategy combinations s_{-i}, t_{-i} *with* $z(s_i, s_{-i}) = z(t_i, t_{-i})$ *it holds that* $u_i(s_i, s_{-i}) = u_i(t_i, t_{-i})$.

This is the traditional way in which consequentialism in dynamic games is modelled. In the sequel, we call a utility function u_i having the property above a *utility function on consequences.*

As an illustration, consider the dynamic game form from Figure 1. Consider the conditional preference relation $\succsim_1$ for you that has the expected utility representation u_1 given by Table 1. Note that $z(($*leave, slam door*$),$ *shout*$) = z(($*leave, slam door*$),$ *do not shout*$)$ but $u_1(($*leave, slam door*$),$ *shout*$) \neq u_1(($*leave, slam door*$),$ *do not shout*$)$. Despite this, it can be shown that $\succsim_1$ is utility-based consequentialist. Indeed, suppose we add the fixed utility 1 to all utilities in the second column, leading to a new utility function v_1. Then, for every belief the expected utility differences between strategies will be the same in u_1 as in v_1, which implies that also v_1 will be an expected utility representation for $\succsim_1$. Moreover, it can be verified that the new utility function v_1 is a utility function on consequences. As such, $\succsim_1$ is utility-based consequentialist.

It is easily seen that every conditional preference relation which is utility-based consequentialist is also preference-based consequentialist. However, as we will see in the following section, the other direction is not always true.

To close this section, let us go back to the example from Figure 1 with the conditional preference relation $\succsim_1$ for you given by the utility function u_1 in Table 1. Suppose we replace the utility 3 by a utility of 6. Then, you prefer (*leave, slam door*) to (*leave, don't slam door*) if you believe that Barbara would start shouting if you were to stay, whereas the ranking would be reversed if you believe that Barbara would not start shouting in this case. Such a conditional preference relation would not be *preference-based consequentialist,* and therefore also not *utility-based consequentialist.*

4 Difference Between the Two Notions

In this section we first present an example where utility-based consequentialism is more restrictive than preference-based consequentialism. We will see that utility-based consequentialism requires the decision maker to hold *additive preference intensities on consequences* – a property that is not required by preference-based consequentialism in this example. We proceed by offering a formal definition of additive preference intensities on consequences and show that, in the absence of weakly dominated strategies, utility-based consequentialism is equivalent to having additive preference intensities on consequences. Moreover, it is

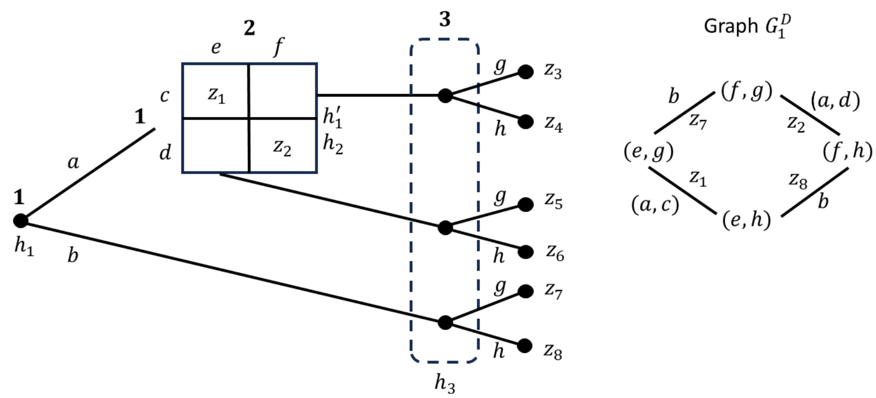

Figure 2: Utility-based consequentalism may be stronger than preference-based consequentialism

	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	z_1	z_3	z_1	z_4
(a,d)	z_5	z_2	z_6	z_2
b	z_7	z_7	z_8	z_8

Table 2: Induced consequences for the dynamic game form in Figure 2

shown that, in the absence of weakly dominated strategies, preference-based consequentialism is equivalent to having additive preference intensities on consequences *for every pair of strategies.*

4.1 Example

We will now present an example where utility-based consequentialism is more restrictive than preference-based consequentialism. Consider the dynamic game form in the left-hand panel of Figure 2. Note that there are three players. The information sets for player 1 are h_1 and h_1', whereas h_2 and h_3 are the unique information sets for players 2 and 3, respectively. The information sets h_1' and h_2 represent a history where players 1 and 2 choose simultaneously. The possible consequences are $z_1, ..., z_8$. At information set h_1', the action pair (c, e) leads to the consequence z_1 whereas (d, f) leads to the consequence z_2. The sets of strategies for the three players are $S_1 = \{(a, c), (a, d), b\}$, $S_2 = \{e, f\}$ and $S_3 = \{g, h\}$, respectively.

We view the dynamic game form from the viewpoint of player 1. Table 2 represents, for every strategy for player 1, and every opponents' strategy combination for players 2 and 3, the induced consequence. It is first shown that every conditional preference relation $\succsim_1$ for player 1 with an expected utility representation and without weakly dominated strategies is preference-based consequentialist. In other words, preference-based consequentialism imposes no additional restrictions.

To see this, consider a conditional preference relation $\succsim_1$ with an expected utility representation u_1 such that no two strategies weakly dominate one another. To show that $\succsim_1$ is preference-based consequentialist, consider four strategies s_1, s_1', t_1, t_1' and two beliefs β_1, β_1' with $\mathbb{P}_{(s_1,\beta_1)} = \mathbb{P}_{(s_1',\beta_1')}$ and $\mathbb{P}_{(t_1,\beta_1)} = \mathbb{P}_{(t_1',\beta_1')}$. We must show that $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1' \succsim_{1,\beta_1'} t_1'$.

As different strategies for player 1 lead to different consequences we must have that $s_1 = s_1'$ and $t_1 = t_1'$. If $s_1 = t_1$ then it trivially holds that $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1' \succsim_{1,\beta_1'} t_1'$. Let us therefore assume that $s_1 \neq t_1$.

Suppose first that $s_1 = (a,c)$ and $t_1 = (a,d)$. Since $\mathbb{P}_{((a,c),\beta_1)} = \mathbb{P}_{((a,c),\beta_1')}$ it follows from Table 2 that $\beta_1(f,g) = \beta_1'(f,g)$ and $\beta_1(f,h) = \beta_1'(f,h)$. Similarly, as $\mathbb{P}_{((a,d),\beta_1)} = \mathbb{P}_{((a,d),\beta_1')}$ it follows that $\beta_1(e,g) = \beta_1'(e,g)$ and $\beta_1(e,h) = \beta_1'(e,h)$. We thus conclude that $\beta_1 = \beta_1'$. But then, it trivially holds that $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1' \succsim_{1,\beta_1'} t_1'$ since $s_1 = s_1'$ and $t_1 = t_1'$.

Suppose next that $s_1 = (a,c)$ and $t_1 = b$. Since $\mathbb{P}_{((a,c),\beta_1)} = \mathbb{P}_{((a,c),\beta_1')}$ we must have that $\beta_1(f,g) = \beta_1'(f,g)$ and $\beta_1(f,h) = \beta_1'(f,h)$. Moreover, as $\mathbb{P}_{(b,\beta_1)} = \mathbb{P}_{(b,\beta_1')}$ it follows that $\beta_1(e,g) + \beta_1(f,g) = \beta_1'(e,g) + \beta_1'(f,g)$ and $\beta_1(e,h) + \beta_1(f,h) = \beta_1'(e,h) + \beta_1'(f,h)$. Altogether, we thus conclude that $\beta_1 = \beta_1'$. But then, it trivially holds that $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1' \succsim_{1,\beta_1'} t_1'$ since $s_1 = s_1'$ and $t_1 = t_1'$.

Suppose finally that $s_1 = (a,d)$ and $t_1 = b$. Since $\mathbb{P}_{((a,d),\beta_1)} = \mathbb{P}_{((a,d),\beta_1')}$ we must have that $\beta_1(e,g) = \beta_1'(e,g)$ and $\beta_1(e,h) = \beta_1'(e,h)$. Moreover, as $\mathbb{P}_{(b,\beta_1)} = \mathbb{P}_{(b,\beta_1')}$ it follows that $\beta_1(e,g) + \beta_1(f,g) = \beta_1'(e,g) + \beta_1'(f,g)$ and $\beta_1(e,h) + \beta_1(f,h) = \beta_1'(e,h) + \beta_1'(f,h)$. Altogether, we thus conclude that $\beta_1 = \beta_1'$. But then, it trivially holds that $s_1 \succsim_{1,\beta_1} t_1$ if and only if $s_1' \succsim_{1,\beta_1'} t_1'$ since $s_1 = s_1'$ and $t_1 = t_1'$. Summarizing, we conclude that $\succsim_1$ is preference-based consequentialist.

We next show that utility-based consequentialism imposes restrictions that are absent under preference-based consequentialism. To see this, consider a conditional preference relation $\succsim_1$ without weakly dominated strategies that is utility-based consequentialist. Then, $\succsim_1$ has an expected utility representation u_1 which is a utility function on consequences. Note from Table 2 that

$$z(b,(e,g)) = z(b,(f,g)) = z_7,\ z((a,d),(f,g)) = z((a,d),(f,h)) = z_2,$$
$$z((a,c),(e,g)) = z((a,c),(e,h)) = z_1 \text{ and } z(b,(e,h)) = z(b,(f,h)) = z_8,$$

which is visualized by the graph G_1^D in the right-hand panel of Figure 2. As u_1 is a utility function on consequences we must have that

$$u_1(b,(e,g)) = u_1(b,(f,g)),\ u_1((a,d),(f,g)) = u_1((a,d),(f,h)),$$
$$u_1((a,c),(e,g)) = u_1((a,c),(e,h)) \text{ and } u_1(b,(e,h)) = u_1(b,(f,h)),$$

which implies that

$$\begin{aligned} & [u_1(b,(f,g)) - u_1((a,d),(f,g))] + [u_1((a,d),(f,h)) - u_1(b,(f,h))] \\ = \quad & [u_1(b,(e,g)) - u_1((a,c),(e,g))] + [u_1((a,c),(e,h)) - u_1(b,(e,h))]\,. \end{aligned} \tag{4.1}$$

Since there are no weakly dominated strategies under $\succsim_1$, it follows from Perea (2023) that the utility differences $v_1(s_1, s_{-1}) - v_1(t_1, s_{-1})$ are unique across all expected utility representations v_1 for $\succsim_1$, up to

	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	-1	0	1	0
(a,d)	0	0	0	0
b	0	0	0	0

Table 3: Non-transitive preferences on consequences for the dynamic game form in Figure 2

a positive multiplicative constant. This means that (4.1) applies to *all* expected utility representations u_1 for $\succsim_1$, and is thus a structural property of the conditional preference relation $\succsim_1$. In fact, it turns out that the restriction in (4.1) characterizes *all* conditional preference relations $\succsim_1$ that are utility-based consequentialist.

But what does (4.1) intuitively mean? In Perea (2023) it is argued that for a conditional preference relation $\succsim_1$ without weakly dominated strategies, the utility difference $v_1(s_1, s_{-1}) - v_1(t_1, s_{-1})$, which is unique up to a positive multiplicative constant, can be interpreted as the *intensity* by which player 1 prefers s_1 to t_1 under the belief that the opponents choose s_{-1}. This intensity will be negative if $v_1(s_1, s_{-1}) < v_1(t_1, s_{-1})$. If we assume consequentialism, as we do in this example, then $v_1(s_1, s_{-1}) - v_1(t_1, s_{-1})$ also represents the *intensity* by which player 1 prefers the consequence $z(s_1, s_{-1})$ to the consequence $z(t_1, s_{-1})$, thus leading to a cardinal interpretation of the utility function.

Consider now the first utility difference in (4.1), which is $u_1(b,(f,g)) - u_1((a,d),(f,g))$. As $z(b,(f,g)) = z_7$ and $z((a,d),(f,g)) = z_2$, the utility difference represents the intensity by which player 1 prefers consequence z_7 to consequence z_2, denoted by $int_{z_7 \succ z_2}$. In a similar way, the second term in (4.1) represents $int_{z_2 \succ z_8}$, the third term represents $int_{z_7 \succ z_1}$, whereas the last term represents $int_{z_1 \succ z_8}$. Put together, (4.1) can be read as

$$int_{z_7 \succ z_2} + int_{z_2 \succ z_8} = int_{z_7 \succ z_1} + int_{z_1 \succ z_8}. \tag{4.2}$$

If we assume that preference intensity between consequences is an additive notion, then both $int_{z_7 \succ z_2} + int_{z_2 \succ z_8}$ and $int_{z_7 \succ z_1} + int_{z_1 \succ z_8}$ represent the intensity by which player 1 prefers consequence z_7 over consequence z_8. As such, condition (4.2), as well as condition (4.1), reflect the assumption that the player's preference intensities on consequences are additive.

Summarizing, we see that utility-based consequentialism requires player 1's preference intensities on consequences to be additive, whereas preference-based consequentialism does not impose such condition in this particular example.

It may even happen in this example that preference-based consequentialism allows for *non-transitive* preferences on consequences. To see this, consider the conditional preference relation $\succsim_1$ given by the expected utility representation u_1 in Table 3. It follows from our findings above that $\succsim_1$ is preference-based consequentialist.

The facts that $u_1(b,(f,g)) = u_1((a,d),(f,g))$ and $u_1((a,d),(f,h)) = u_1(b,(f,h))$ seem to suggest that player 1 is indifferent between consequences z_7 and z_2, and is indifferent between z_2 and z_8. On the other hand, $u_1(b,(e,g)) > u_1((a,c),(e,g))$ and $u_1((a,c),(e,h)) > u_1(b,(e,h))$ seem to indicate that player 1 prefers z_7 to z_1, and prefers z_1 to z_8. This can only be if player 1's preferences over consequences are non-transitive.

4.2 Additive Preference Intensities on Consequences

Based on the example above we will now give a formal expression of *additive preference intensities on consequences,* which is implied by utility-based consequentialism. To this purpose we need the following piece of notation: For a strategy s_i, a pair of opponents' strategy combinations s_{-i}, t_{-i} and a consequence z we write $s_{-i} \overset{s_i,z}{-} t_{-i}$ if $z(s_i, s_{-i}) = z(s_i, t_{-i}) = z$.

Definition 4.1 (Additive preference intensities on consequences) *Consider a conditional preference relation $\succsim_i$ with an expected utility representation u_i and without weakly dominated strategies. Then, $\succsim_i$ induces **additive preference intensities on consequences** if for every two opponents' strategy combinations s^*_{-i}, t^*_{-i}, and every two paths*

$$s^*_{-i} \overset{s_i^1,z^1}{-} s^2_{-i} \overset{s_i^2,z^2}{-} s^3_{-i} \dots \overset{s_i^{K-1},z^{K-1}}{-} s^K_{-i} \overset{s_i^K,z^K}{-} t^*_{-i}$$

and

$$s^*_{-i} \overset{t_i^1,y^1}{-} t^2_{-i} \overset{t_i^2,y^2}{-} t^3_{-i} \dots \overset{t_i^{L-1},y^{L-1}}{-} t^L_{-i} \overset{t_i^L,y^L}{-} t^*_{-i}$$

from s^*_{-i} to t^*_{-i} it holds that

$$\begin{aligned} & \left[u_i(s_i^1, s^2_{-i}) - u_i(s_i^2, s^2_{-i})\right] + \left[u_i(s_i^2, s^3_{-i}) - u_i(s_i^3, s^3_{-i})\right] + \dots \\ & \dots + \left[u_i(s_i^{K-1}, s^K_{-i}) - u_i(s_i^K, s^K_{-i})\right] + \left[u_i(s_i^K, t^*_{-i}) - u_i(t_i^L, t^*_{-i})\right] \\ = \; & \left[u_i(s_i^1, s^*_{-i}) - u_i(t_i^1, s^*_{-i})\right] + \left[u_i(t_i^1, t^2_{-i}) - u_i(t_i^2, t^2_{-i})\right] + \\ & + \left[u_i(t_i^2, t^3_{-i}) - u_i(t_i^3, t^3_{-i})\right] + \dots + \left[u_i(t_i^{L-1}, t^L_{-i}) - u_i(t_i^L, t^L_{-i})\right]. \end{aligned}$$

As there are no weakly dominated strategies under $\succsim_i$, it follows by Perea (2023) that the sums of the utility differences on the left-hand side and right-hand side are unique up to a (common) positive multiplicative constant. Therefore, the equality is a structural property of $\succsim_i$ that holds for *all* expected utility representations u_i for $\succsim_i$.

Note that the sum of the utility differences on the left-hand side represents

$$int_{z^1 \succ z^2} + int_{z^2 \succ z^3} + \dots + int_{z^{K-1} \succ z^K} + int_{z^K \succ y^L} \tag{4.3}$$

whereas the sum of the utility differences on the right-hand side amounts to

$$int_{z^1 \succ y^1} + int_{y^1 \succ y^2} + int_{y^2 \succ y^3} + \dots + int_{y^{L-1} \succ y^L}. \tag{4.4}$$

The condition in the definition thus states that the sums of the preference intensities in (4.3) and (4.4) must be equal. As, under additivity, both sums represent the intensity by which player 1 prefers consequence z^1 to consequence y^L, the condition in the definition reflects the assumption that the player's preference intensities on consequences are additive.

	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	z_1	z_3	z_1	z_4
(a,d)	z_5	z_2	z_6	z_2
b	z_7	z_7	z_8	z_8

u_1	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	x_1	x_4	x_7	x_{10}
(a,d)	x_2	x_5	x_8	x_{11}
b	x_3	x_6	x_9	x_{12}

Table 4: Consequence mapping and expected utility representation in the dynamic game form of Figure 2

4.3 Characterization of Utility-Based Consequentialism

It turns out that the condition of additive preference intensities on consequences characterizes precisely those conditional preference relations that are utility-based consequentialist.

Theorem 4.1 (Characterization of utility-based consequentialism) *Consider a dynamic game form D, a player i and a conditional preference relation $\succsim_i$ for player i that has an expected utility representation and under which there are no weakly dominated strategies. Then, $\succsim_i$ is utility-based consequentialist, if and only if, $\succsim_i$ induces additive preference intensities on consequences.*

It is relatively easy to show that under the conditions in the theorem, utility-based consequentialism implies that the conditional preference relation induces additive preference intensities on consequences. We basically follow the steps we have performed in the example of Figure 2 above.

Showing the other direction is more difficult: Under the conditions in the theorem, and assuming that $\succsim_i$ induces additive preference intensities on consequences, we explicitly show how to transform an arbitrary expected utility representation u_i into a new expected utility representation v_i that is a utility function on consequences. We will now illustrate this direction of the proof by means of the example of Figure 2.

We will again view the situation from player 1's perspective. The induced consequences are repeated in the left-hand panel of Table 4. Suppose that the conditional preference relation $\succsim_1$ is given by the expected utility representation u_1 in the right-hand panel of Table 4, where $x_1, ..., x_{12}$ represent the 12 utilities. Assume that the utility function u_1 is such that $\succsim_1$ induces additive preference intensities on consequences, and that there are no weakly dominated strategies for player 1.

We now transform u_1, in a step-by-step fashion, into a new expected utility representation v_1 that is a utility function on consequences, as follows. We keep the utilities x_1, x_2 and x_3 in column (e,g) as they are.

We then move to column (f,g). Note that $z(b,(e,g)) = z(b,(f,g))$. At column (f,g) we therefore add a constant utility $x_3 - x_6$ to the entries in that column such that $v_1(b,(e,g)) = v_1(b,(f,g))$.

Also, $z((a,c),(e,g)) = z((a,c),(e,h))$. Similarly, we then add a constant utility $x_1 - x_7$ to the entries in column (e,h) such that $v_1((a,c),(e,g)) = v_1((a,c),(e,h))$. This leads to the utility function in the left-hand panel of Table 5 . Here, the numbers y_4, y_5, x_3, x_1, y_8 and y_9 in the second and third column denote the new utilities for v_1 in those columns.

Finally, we move to the remaining column (f,h). Note that $z((a,d),(f,g)) = z((a,d),(f,h))$ and $z(b,(e,h)) = z(b,(f,h))$. At column (f,h) we add a constant utility $y_5 - x_{11}$ to the entries in that column such that $v_1((a,d),(f,g)) = v_1((a,d),(f,h))$. This leads to the utility function v_1 in the right-hand

	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	x_1	y_4	x_1	x_{10}
(a,d)	x_2	y_5	y_8	x_{11}
b	x_3	x_3	y_9	x_{12}

	(e,g)	(f,g)	(e,h)	(f,h)
(a,c)	x_1	y_4	x_1	y_{10}
(a,d)	x_2	y_5	y_8	y_5
b	x_3	x_3	y_9	y_{12}

Table 5: Construction of utility function v_1 in the dynamic game form of Figure 2

panel of Table 5. As v_1 has been obtained from u_1 by adding a constant utility to each of the columns, it follows that v_1 is also an expected utility representation of $\succsim_1$. The procedure we have used here is called the *utility transformation procedure,* and is described formally in the appendix.

We will now show that v_1 is a utility function on consequences, by proving that $y_9 = y_{12}$. Our construction above guarantees that

$$\begin{aligned} v_1(b,(e,g)) = v_1(b,(f,g)),\ v_1((a,c),(e,g)) = v_1((a,c),(e,h)) \\ \text{and } v_1((a,d),(f,g)) = v_1((a,d),(f,h)). \end{aligned} \tag{4.5}$$

Consider the graph G_1^D in the right-hand panel of Figure 2. Note that this graph contains two alternative paths from (e,g) to (f,h). As $\succsim_1$ induces additive preference intensities on consequences, we conclude that

$$\begin{aligned} & [v_1(b,(f,g)) - v_1((a,d),(f,g))] + [v_1((a,d),(f,h)) - v_1(b,(f,h))] \\ = \ & [v_1(b,(e,g)) - v_1((a,c),(e,g))] + [v_1((a,c),(e,h)) - v_1(b,(e,h))]\,. \end{aligned} \tag{4.6}$$

By combining (4.5) and (4.6) we conclude that $v_1(b,(f,h)) = v_1(b,(e,h))$, and hence $y_9 = y_{12}$. Therefore, v_1 is a utility function on consequences. As v_1 is an expected utility representation for $\succsim_1$, it follows that $\succsim_1$ is utility-based consequentialist.

4.4 Characterization of Preference-Based Consequentialism

Above we have seen that utility-based consequentialism can be characterized by requiring that the induced preference intensities on consequences are additive. This raises the question: How does preference-based consequentialism relate to additive preference intensities on consequences? The following result shows that this weaker version of consequentialism is equivalent to demanding that every *pair of strategies* induces additive preference intensities on consequences.

To formally state this result we need the following piece of notation. For a given conditional preference relation $\succsim_i$ and pair of strategies $\{s_i,t_i\}$, we denote by $\succsim_i^{\{s_i,t_i\}}$ the restriction of $\succsim_i$ to the strategies s_i and t_i. That is, $\succsim_i^{\{s_i,t_i\}}$ ranks, for every belief, only the strategies s_i and t_i, and for every belief β_i we have that $s_i \succsim_{i,\beta_i}^{\{s_i,t_i\}} t_i$ if and only if $s_i \succsim_{i,\beta_i} t_i$ and $t_i \succsim_{i,\beta_i}^{\{s_i,t_i\}} s_i$ if and only if $t_i \succsim_{i,\beta_i} s_i$.

Theorem 4.2 (Characterization of preference-based consequentialism) *Consider a dynamic game form D, a player i and a conditional preference relation $\succsim_i$ for player i that has an expected utility representation and under which there are no weakly dominated strategies. Then, $\succsim_i$ is preference-based consequentialist, if and only if, for every pair of strategies s_i, t_i the restricted conditional preference relation $\succsim_i^{\{s_i,t_i\}}$ induces additive preference intensities on consequences.*

In view of the Theorems 4.1 and 4.2, the difference between utility-based and preference-based consequentialism can be characterized by the induced preference intensities on consequences: Utility-based consequentialism requires these preference intensities to be additive for the set of *all* strategies, whereas preference-based consequentialism only demands this property for every *pair* of strategies in isolation.

5 When the Two Notions are Equivalent

As the example in Figure 2 has shown, there are dynamic game forms where preference-based and utility-based consequentialism are different. The reason is that utility-based consequentialism implies additive preference intensities over consequences, whereas preference-based consequentialism only implies this property for every pair of strategies. The example in Figure 2 shows that the first condition may be more demanding than the second. It may thus be argued that for these scenarios, the notion of utility-based consequentialism imposes more than what is required by the original idea of consequentialism.

We will now provide sufficient conditions under which the two notions of consequentialism are equivalent.

Theorem 5.1 (Equivalence) *Consider a dynamic game form D and a player i such that either (i) player i only has two strategies, (ii) D has observed past choices, or (iii) D only has two players and satisfies perfect recall. Moreover, consider a conditional preference relation $\succsim_i$ for player i without weakly dominated strategies that has an expected utility representation. Then, $\succsim_i$ is preference-based consequentialist, if and only if, $\succsim_i$ is utility-based consequentialist.*

Note that the example from the previous section, where the two notions of consequentialism are not equivalent, violates the conditions (i), (ii) and (iii) above. Indeed, the dynamic game form in the example has more than two strategies for player 1, violates observed past choices and has more than two players. To see that it violates observed past choices, note that player 3, at his information set h_3, does not perfectly observe what players 1 and 2 have chosen in the past.

We will now provide a sketch of the proof. The easy direction is to show that utility-based consequentialism implies preference-based consequentialism. The other direction is more challenging: We must show that, under the conditions of the theorem, every conditional preference relation $\succsim_i$ that is prefence-based consequentialist is also utility-based consequentialist. We do so by transforming the utility function u_i that represents $\succsim_i$ into an expected utility representation v_i that is a utility function on consequences, in the same way as we did in the proof of Theorem 4.1.

We illustrate this direction of the proof by a new example. Consider the dynamic game form D between player 1 and player 2 in the left-hand panel of Figure 3, with the associated consequence mapping in the left-hand panel of Table 6. We will view the situation from player 1's perspective. Suppose that the conditional

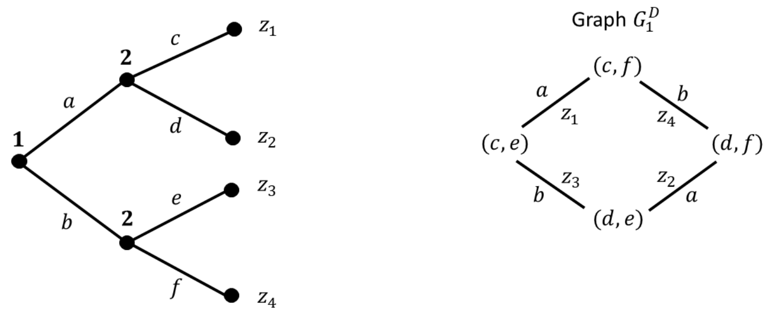

Figure 3: Proof sketch of Theorem 5.1

	(c,e)	(c,f)	(d,e)	(d,f)
a	z_1	z_1	z_2	z_2
b	z_3	z_4	z_3	z_4

u_1	(c,e)	(c,f)	(d,e)	(d,f)
a	x_1	x_3	x_5	x_7
b	x_2	x_4	x_6	x_8

Table 6: Consequence mapping and expected utility representation in the dynamic game form of Figure 3

	(c,e)	(c,f)	(d,e)	(d,f)
a	x_1	x_1	y_5	x_7
b	x_2	y_4	x_2	x_8

	(c,e)	(c,f)	(d,e)	(d,f)
a	x_1	x_1	y_5	y_7
b	x_2	y_4	x_2	y_4

Table 7: Construction of utility function v_1 in the dynamic game form of Figure 3

preference relation $\succsim_1$ is given by the expected utility representation u_1 in the right-hand panel of Table 6. Assume that the utility function u_1 is such that $\succsim_1$ is preference-based consequentialist, and that there are no weakly dominated strategies for player 1.

We now transform u_1, in a step-by-step fashion, into a new expected utility representation v_1 that is a utility function on consequences, as follows. We keep the utilities x_1 and x_2 in column (c,e) as they are.

We then move to column (c,f). Note that $z(a,(c,e)) = z(a,(c,f))$. At column (c,f) we therefore add a constant utility $x_1 - x_3$ to the entries in that column such that $v_1(a,(c,e)) = v_1(a,(c,f))$.

Also, $z(b,(c,e)) = z(b,(d,e))$. Similarly, we then add a constant utility $x_2 - x_6$ to the entries in column (d,e) such that $v_1(b,(c,e)) = v_1(b,(d,e))$. This leads to the utility function in the left-hand panel of Table 7. Here, the numbers x_1, y_4, y_5 and x_2 in the second and third column denote the new utilities for v_1 in those columns.

Finally, we move to the remaining column (d,f). Note that $z(a,(d,e)) = z(a,(d,f))$ and $z(b,(c,f)) = z(b,(d,f))$. At column (d,f) we add a constant utility $y_4 - x_8$ to the entries in that column such that $v_1(b,(c,f)) = v_1(b,(d,f))$. This leads to the utility function v_1 in the right-hand panel of Table 7.

We will now show that v_1 is a utility function on consequences, by proving that $y_5 = y_7$. Our construction above guarantees that

$$v_1(a,(c,e)) = v_1(a,(c,f)),\ v_1(b,(c,e)) = v_1(b,(d,e)) \text{ and } v_1(b,(c,f)) = v_1(b,(d,f)). \tag{5.1}$$

Consider the beliefs $\beta_1 := \frac{1}{2}[(c,e)] + \frac{1}{2}[(d,f)]$ and $\beta_1' := \frac{1}{2}[(c,f)] + \frac{1}{2}[(d,e)]$, where $[s_2]$ is the probability distribution that assigns probability 1 to player 2's strategy s_2. Then, we conclude from the consequence mapping in Table 6 that

$$\mathbb{P}_{(a,\beta_1)} = \mathbb{P}_{(a,\beta_1')} = \tfrac{1}{2}[z_1] + \tfrac{1}{2}[z_2] \text{ and } \mathbb{P}_{(b,\beta_1)} = \mathbb{P}_{(b,\beta_1')} = \tfrac{1}{2}[z_3] + \tfrac{1}{2}[z_4].$$

Since $\succsim_1$ is assumed to be preference-based consequentialist, we know that $a \succsim_{1,\beta_1} b$ if and only if $a \succsim_{1,\beta_1'} b$. As there are no weakly dominated strategies for player 1, it can be shown that this implies that $v_1(a,\beta_1) - v_1(b,\beta_1) = v_1(a,\beta_1') - v_1(b,\beta_1')$, which means that

$$\begin{aligned} & \tfrac{1}{2}v_1(a,(c,e)) + \tfrac{1}{2}v_1(a,(d,f)) - \tfrac{1}{2}v_1(b,(c,e)) - \tfrac{1}{2}v_1(b,(d,f)) \\ = \quad & \tfrac{1}{2}v_1(a,(c,f)) + \tfrac{1}{2}v_1(a,(d,e)) - \tfrac{1}{2}v_1(b,(c,f)) - \tfrac{1}{2}v_1(b,(d,e)). \end{aligned} \tag{5.2}$$

By combining (5.1) and (5.2) it then follows that $v_1(a,(d,f)) = v_1(a,(d,e))$. That is, $y_5 = y_7$, which was to show. We thus obtain an expected utility representation v_1 for $\succsim_1$ which is a utility function on consequences. As such, $\succsim_1$ is utility-based consequentialist.

In the proof of Theorem 5.1 the construction of the new utility function v_i proceeds along the same lines. The construction is based on a graph G_i^D where two columns (opponents' strategy combinations) s_{-i} and t_{-i} are "connected" by a strategy s_i if s_{-i} and t_{-i} only differ at one information set[2] and $z(s_i, s_{-i}) = z(s_i, t_{-i})$. Such a connection means that the utilities at s_{-i} and t_{-i} are interrelated, since we must make sure that $v_i(s_i, s_{-i}) = v_i(s_i, t_{-i})$. For every connected component in the graph G_i^D we start by copying the utilities of u_i at a distinguished column s_{-i}^0, and step by step we construct the new utilities of v_i at the other columns by following sequences of connected columns, in the same way as we have done for the example above. The graph G_1^D for the example above can be found in the right-hand panel of Figure 3. The label a at the edge between (c, e) and (c, f) indicates that $z(a, (c, e)) = z(a, (c, f))$, and similarly for the other edges.

Showing that v_i is a utility function on consequences only poses problems if there is a column s_{-i} that can be reached through two different paths of connected columns from s_{-i}^0, thus yielding a cycle. This was the case in the graph G_1^D above, since the column (d, f) could be reached through the path $(c, e) \to (c, f) \to (d, f)$ but also through the path $(c, e) \to (d, e) \to (d, f)$, yielding the cycle $(c, e) \to (c, f) \to (d, f) \to (d, e) \to (c, e)$. In the proof of Theorem 5.1 we show that the conditions (i), (ii) or (iii) on the dynamic game form in the theorem guarantee that there are at most two strategies, s_i and t_i, that connect all the columns in the cycle. Similarly to the example above, such a cycle then induces two beliefs β_i and β_i' such that $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$ and $\mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i,\beta_i')}$. As $\succsim_i$ satisfies preference-based consequentialism, we can derive equalities like (5.2) to show that v_i is a utility function on consequences.

6 Concluding Remarks

Observable versus non-onbservable consequences. Throughout this paper we have assumed that the consequences are *observable*, by identifying these with sequences of *realized* action profiles that lead to a terminal history in the game. Our notion of consequentialism relies on this assumption. Alternatively, one may extend the set of consequences to collections of realized and *non-realized* (hypothetical) actions. As an illustration, consider the example from Figure 1 in the introduction. We argued that it is reasonable not to be a consequentialist in this game, by preferring to slam the door if you believe that Barbara would conterfactually have started to shout if you had decided to stay, and preferring to leave calmly if you believe that Barbara would have remained friendly in this case.

Some people may object to this conclusion, because slamming the door while believing that Barbara would have started to shout may be viewed as a different (partially non-observable) consequence than slamming door while believing that she would not have started to shout. Indeed, by choosing this avenue the set of consequences could be extended to

(*stay, shout*), (*stay, don't shout*), ((*leave, slam door*), *shout*), ((*leave, slam door*), *don't shout*),
((*leave, don't slam door*), *shout*), ((*leave, don't slam door*), *don't shout*),

[2]More precisely, if $(s_{-i}) = (s_j)_{j \neq i}$ and $t_{-i} = (t_j)_{j \neq i}$ then there is an opponent j and an information set h_j such that s_j and t_j only differ at h_j and the information sets that follow, whereas $s_k = t_k$ for all other opponents k. See the appendix for more details.

where the last four consequences are partially unobservable because they involve a non-realized, hypothetical action by Barbara.

Still, the preferences described above could be summarized by

$$((\textit{leave, slam door}),\ \textit{shout}) \succ ((\textit{leave, don't slam door}),\ \textit{shout})$$

and

$$((\textit{leave, don't slam door}),\ \textit{don't shout}) \succ ((\textit{leave, slam door}),\ \textit{don't shout}).$$

Hence, according to the new, extended set of partially unobservable consequences, the preferences above would be consistent with consequentialism.

As such, the notion of consequentialism crucially depends on how one defines the set of consequences. In this paper, we have made the choice of identifying consequences with sequences of realized actions. In particular, we restrict to consequences that are fully observable.

Writing down utilities at consequences may imply more than consequentialism. The analysis in this paper has shown that writing down utilities at the terminal nodes in a dynamic game, resulting in utility-based consequentialism, may imply conditions that go beyond preference-based consequentialism. Indeed, we have characterized utility-based consequentialism by the condition that the induced preference intensities on consequences are additive, and the example from Figure 2 indicates that this condition need not follow from preference-based consequentialism. For such situations it may thus be argued that utility-based consequentialism is more restrictive than the original idea of consequentialism.

Possible extensions of our results. We have identified conditions on dynamic game forms under which preference-based consequentialism is equivalent to utility-based consequentialism, and where the condition of additive preference intensities on consequences is thus implied by preference-based consequentialism alone. An open question is whether these conditions on the dynamic game form can be sharpened to conditions that are both sufficient *and necessary* for the equivalence. That is, if the conditions are violated, then we can find a conditional preference relation that is preference-based, but not utility-based, consequentialist.

Also, the main results in the paper rely on the assumption that there are no weakly dominated strategies under the conditional preference relation we consider. It is currently unclear whether, and if so how, these results can be extended to situations that allow for weakly dominated strategies.

7 Appendix

7.1 Definitions from Graph Theory

An *undirected graph* $G = (N, E)$ consists of a set of nodes N, and a set of edges E, where every edge $e \in E$ is an unordered pair $(n, n') \in N \times N$ with $n \neq n'$. A graph $G' = (N', E')$ is a *subgraph* of $G = (N, E)$ if $N' \subseteq N$, $E' \subseteq E$ and every edge $(n, n') \in E'$ is such that $n, n' \in N'$.

In a graph $G = (N, E)$, a *path* from $n \in N$ to $n' \in N$ is a sequence $(n^0, n^1, ..., n^K)$ with $n^0 = n$ and $n^K = n'$ such that $(n^k, n^{k+1}) \in E$ for every $k \in \{0, ..., K-1\}$. A *cycle* is a path $(n^0, n^1, ..., n^K)$ where $n^K = n^0$.

A subgraph $C = (N', E')$ of $G = (N, E)$ is a *connected component* of G if (i) $E' = \{(n, n') \in E \mid n, n' \in N'\}$, (ii) for every two nodes $n, n' \in N'$ there is a path from n to n' in G, and (iii) for every $n \in N'$, $n' \in N \backslash N'$ there is no path from n to n' in G.

A graph $T = (N, E)$ is a *tree* if there is some $n^0 \in N$ such that for every $n \in N \backslash \{n^0\}$ there is a unique path in T from n^0 to n. In this case, we call T a tree with *root* n^0. A subgraph $T = (N', E')$ of $G = (N, E)$ is a *spanning tree for* G if $N' = N$ and T is a tree. For a given graph G, it is well-known that for every connected component C of G there is a spanning tree for C.

7.2 Preparatory Results

To prove the theorems in this paper we need some preparatory results.

Lemma 7.1 (Implication of preference-based consequentialism) *Consider a conditional preference relation $\succsim_i$ that is preference-based consequentialist, two strategies s_i, t_i that do not weakly dominate one another under $\succsim_i$, and an expected utility representation u_i for $\succsim_i$. Then, for all beliefs $\beta_i, \beta_i' \in \Delta(S_{-i})$ such that $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$ and $\mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i,\beta_i')}$ we have that $u_i(s_i, \beta_i) - u(t_i, \beta_i) = u(s_i, \beta_i') - u_i(t_i, \beta_i')$.*

Proof. Since s_i and t_i do not weakly dominate one another, it follows from Perea (2023) that there is a belief β_i^* with $\beta_i^*(s_{-i}) > 0$ for all $s_{-i} \in S_{-i}$ such that $s_i \sim_{i,\beta_i^*} t_i$. We can choose $\varepsilon > 0$ small enough such that $\beta_i^{''} := \beta_i^* + \varepsilon(\beta_i - \beta_i')$ is a belief. We show that $\mathbb{P}_{(s_i,\beta_i'')} = \mathbb{P}_{(s_i,\beta_i^*)}$ and $\mathbb{P}_{(t_i,\beta_i'')} = \mathbb{P}_{(t_i,\beta_i^*)}$.

Indeed, for every consequence z we have that

$$\mathbb{P}_{(s_i,\beta_i'')}(z) = \mathbb{P}_{(s_i,\beta_i^*)}(z) + \varepsilon(\mathbb{P}_{(s_i,\beta_i)}(z) - \mathbb{P}_{(s_i,\beta_i')}(z)) = \mathbb{P}_{(s_i,\beta_i^*)}(z),$$

since $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$. In a similar way it can be shown that $\mathbb{P}_{(t_i,\beta_i'')}(z) = \mathbb{P}_{(t_i,\beta_i^*)}(z)$ for every consequence z.

Since $s_i \sim_{i,\beta_i^*} t_i$ and $\succsim_i$ is preference-based consequentialist, it follows that $s_i \sim_{i,\beta_i''} t_i$ also. As u_i is an expected utility representation for $\succsim_i$ we know that $u_i(s_i, \beta_i^*) = u_i(t_i, \beta_i^*)$ and $u_i(s_i, \beta_i'') = u_i(t_i, \beta_i'')$. Hence,

$$\begin{aligned} 0 &= u_i(s_i, \beta_i'') - u_i(t_i, \beta_i'') \\ &= (u_i(s_i, \beta_i^*) - u_i(t_i, \beta_i^*)) + \varepsilon((u_i(s_i, \beta_i) - u_i(t_i, \beta_i)) - (u_i(s_i, \beta_i') - u_i(t_i, \beta_i'))) \\ &= \varepsilon((u_i(s_i, \beta_i) - u_i(t_i, \beta_i)) - (u_i(s_i, \beta_i') - u_i(t_i, \beta_i'))), \end{aligned}$$

where the second equality follows from the definition of β_i'', and the third equality follows from the fact that $u_i(s_i, \beta_i^*) = u_i(t_i, \beta_i^*)$. We thus conclude that $u_i(s_i, \beta_i) - u(t_i, \beta_i) = u(s_i, \beta_i') - u_i(t_i, \beta_i')$. This completes the proof. ■

Lemma 7.2 (Constant utility carries over) *Consider a conditional preference relation $\succsim_i$ that is preference-based consequentialist, two strategies s_i, t_i that do not weakly dominate one another, and an expected utility representation u_i for $\succsim_i$. Take two opponents' strategy combinations s_{-i}, t_{-i} with $z(s_i, s_{-i}) = z(s_{-i}, t_{-i})$, $z(t_i, s_{-i}) = z(t_i, t_{-i})$ and $u_i(s_i, s_{-i}) = u_i(s_i, t_{-i})$. Then, $u_i(t_i, s_{-i}) = u_i(t_i, t_{-i})$.*

Proof. If we define the beliefs $\beta_i := [s_{-i}]$ and $\beta_i' := [t_{-i}]$ it follows that $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$ and $\mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i,\beta_i')}$. By Lemma 7.1 we conclude that $u_i(s_i, \beta_i) - u(t_i, \beta_i) = u(s_i, \beta_i') - u_i(t_i, \beta_i')$, and hence $u_i(s_i, s_{-i}) - u(t_i, s_{-i}) = u(s_i, t_{-i}) - u_i(t_i, t_{-i})$. Since $u_i(s_i, s_{-i}) = u_i(s_i, t_{-i})$ it follows that $u_i(t_i, s_{-i}) = u_i(t_i, t_{-i})$, which completes the proof. ■

For the following result we need some additional definitions. We say that two strategies $s_i, t_i \in S_i$ are *minimally different* if there is an information set $h_i \in H_i(s_i) \cap H_i(t_i)$ such that (i) $s_i(h_i) \neq t_i(h_i)$, and (ii) $s_i(h_i') = t_i(h_i')$ for all $h_i' \in (H_i(s_i) \cap H_i(t_i)) \backslash \{h_i\}$. In this case, we call s_i, t_i minimally different *at* h_i. Two strategy combinations $s_{-i} = (s_j)_{j \neq i}$ and $t_{-i} = (t_j)_{j \neq i}$ in S_{-i} are called *minimally different* if there is some $j \neq i$ such that s_j, t_j are minimally different at some $h_j \in H_j(s_j) \cap H_j(t_j)$, and $s_k = t_k$ for all $k \neq i, j$. In this case, we say that s_{-i}, t_{-i} are minimally different at h_j.

Lemma 7.3 (Equal consequences) *Consider a dynamic game form D with two players, i and j, that satisfies perfect recall. Let the strategies s_j, t_j be minimally different at the information set $h_j \in H_j(s_j) \cap H_j(t_j)$. Then, for every strategy s_i we have that $z(s_i, s_j) = z(s_i, t_j)$ if and only if $s_i \notin S_i(h_j)$.*

Proof. **(a)** Suppose first that $z(s_i, s_j) = z(s_i, t_j)$. Then, $(s_i, s_j) \notin S(h_j)$. By perfect recall we have that $S(h_j) = S_i(h_j) \times S_j(h_j)$. Since $s_j \in S_j(h_j)$ we conclude that $s_i \notin S_i(h_j)$.

(b) Suppose next that $s_i \notin S_i(h_j)$. Then, by definition, $(s_i, s_j) \notin S(h_j)$. But then, $z(s_i, s_j) = z(s_i, t_j)$. The proof is hereby complete. ■

To formally express the condition of *two strategies per connected component*, which plays an important role in the proof of Theorem 5.1, we need the following definition. For a dynamic game form D with distinguished player i, consider the undirected graph $G_i^D = (N, E)$ where (i) the set of nodes N is the set of all strategy combinations in S_{-i}, and (ii) the set of edges E contains exactly those pairs $(s_{-i}, t_{-i}) \in N \times N$ where s_{-i}, t_{-i} are minimally different and there is some strategy $s_i \in S_i$ with $z(s_i, s_{-i}) = z(s_i, t_{-i})$. In this case, we also denote this edge by $s_{-i} \overset{s_i,z}{-} t_{-i}$, where $z = z(s_i, s_{-i}) = z(s_i, t_{-i})$.

Definition 7.1 (Two strategies per connected component) *The graph G_i^D satisfies* ***two strategies per connected component*** *if for every connected component C there are two strategies $s_i, t_i \in S_i$ such that for every edge (s_{-i}, t_{-i}) in C either $z(s_i, s_{-i}) = z(s_i, t_{-i})$ or $z(t_i, s_{-i}) = z(t_i, t_{-i})$.*

The following result states that the condition of two strategies per connected component is always satisfied under the conditions on the dynamic game form in Theorem 5.1.

Lemma 7.4 (Two strategies per connected component) *Consider a dynamic game form D and a player i such that either (i) player i only has two strategies, or (ii) there are observed past choices, or (iii) there are only two players and perfect recall is satisfied. Then, the induced graph G_i^D satisfies two strategies per connected component.*

Proof. **(i)** If player i only has two strategies, it trivially follows that G_i^D satisfies two strategies per connected component.

(ii) Assume next that the dynamic game D is with observed past choices. Let H_i^{first} be the collection of information sets in H_i that are not preceded by any other information in H_i. For every $h_i \in H_i^{first}$ select two different actions $a_i(h_i), b_i(h_i) \in A_i(h_i)$. Let s_i^* be a strategy with $s_i^*(h_i) = a_i(h_i)$ for all $h_i \in H_i^{first}$, and t_i^* a strategy with $t_i^*(h_i) = b_i(h_i)$ for all $h_i \in H_i^{first}$.

Now, consider an edge (s_{-i}, t_{-i}) in G_i^D with $(s_{-i}) = (s_j)_{j \neq i}$ and $(t_{-i}) = (t_j)_{j \neq i}$. Then, s_{-i}, t_{-i} are minimally different at some $h_j \in H_j(s_j) \cap H_j(t_j)$, and there is some strategy s_i with $z(s_i, s_{-i}) = z(s_i, t_{-i})$. We distinguish two cases: (1) $h_j \in H_j(s_{-i}) \cap H_j(t_{-i})$, and (2) $h_j \notin H_j(s_{-i}) \cap H_j(t_{-i})$.

Case 1. Suppose that $h_j \in H_j(s_{-i}) \cap H_j(t_{-i})$. As $z(s_i, s_{-i}) = z(s_i, t_{-i})$ and (s_{-i}, t_{-i}) are minimally different at h_j, it must be that $(s_i, s_{-i}) \notin S(h_j)$. Since the game is with observed past choices we know that $S(h_j) = S_i(h_j) \times S_{-i}(h_j)$. Note that $s_{-i} \in S_{-i}(h_j)$ as $h_j \in H_j(s_{-i})$. But then, $(s_i, s_{-i}) \notin S(h_j)$ implies that $s_i \notin S_i(h_j)$. This can only be if h_j is preceded by some $h_i \in H_i^{first}$.

As the game is with observed past choices, there is a unique action $a_i^*(h_i) \in A_i(h_i)$ that leads to h_j. By construction, either $s_i^*(h_i) \neq a_i^*(h_i)$ or $t_i^*(h_i) \neq a_i^*(h_i)$. This means that either $(s_i^*, s_{-i}) \notin S(h_j)$ or $(t_i^*, s_{-i}) \notin S(h_j)$. As s_{-i}, t_{-i} are minimally different at h_j we conclude that either $z(s_i^*, s_{-i}) = z(s_i^*, t_{-i})$ or $z(t_i^*, s_{-i}) = z(t_i^*, t_{-i})$.

Case 2. Suppose that $h_j \notin H_j(s_{-i}) \cap H_j(t_{-i})$. Since s_{-i} and t_{-i} only differ at h_j and afterwards, it follows that $h_j \notin H_j(s_{-i})$, which implies that $(s_i, s_{-i}) \notin S(h_j)$ for every strategy s_i. But then, $z(s_i, s_{-i}) = z(s_i, t_{-i})$ for every strategy s_i. In particular, $z(s_i^*, s_{-i}) = z(s_i^*, t_{-i})$.

In view of Cases 1 and 2, two strategies per connected component holds.

(iii) Suppose finally that the dynamic game form D is with two players, i and j, and that it satisfies perfect recall. Take a connected component C in the induced graph G_i^D, and let

$$H_j(C) := \{h_j \in H_j \mid \text{there is an edge } (s_j, t_j) \text{ in } C \text{ such that } s_j, t_j \text{ minimally different at } h_j\}.$$

Let $H_j^{first}(C)$ be the collection of information sets in $H_j(C)$ that are not preceded by any other information set in $H_j(C)$.

Claim 1. For every $h_j, h_j' \in H_j^{first}(C)$ there is a strategy s_j in C with $s_j \in S_j(h_j) \cap S_j(h_j')$.

Proof of claim 1. Take two different $h_j, h_j' \in H_j^{first}(C)$. Then, by definition, there are edges (s_j, t_j) and (s_j', t_j') in C such that s_j, t_j are minimally different at h_j and s_j', t_j' are minimally different at h_j'. In particular, $s_j \in S_j(h_j)$ and $t_j' \in S_j(h_j')$. Since $t_j, s_j' \in C$, there is a path $(s_j^1, ..., s_j^K)$ in C from t_j to s_j'. Hence, there are information sets $h_j^1, ..., h_j^{K-1} \in H_j(C)$, such that for every $k \in \{1, ..., K-1\}$ the strategies s_j^k, s_j^{k+1} are minimally different at $h_j^k \in H_j(C)$. This implies that s_j^1 and s_j^K only differ at information sets in $H_j(C)$. Recall that $s_j^1 = t_j$ and $s_j^K = s_j'$. As s_j, t_j are minimally different at $h_j \in H_j(C)$ and s_j', t_j' are minimally different at $h_j' \in H_j(C)$, it follows that s_j and t_j' only differ at information sets in $H_j(C)$.

Hence, s_j and t'_j coincide at information sets in H_j that precede information sets in $H_j^{first}(C)$. As $h'_j \in H_j^{first}(C)$, this implies that s_j and t'_j coincide at information sets in H_j that precede h'_j. Since $t'_j \in S_j(h'_j)$ we conclude that $s_j \in S_j(h'_j)$ as well. Recall that $s_j \in S_j(h_j)$. Therefore, s_j is in C and $s_j \in S_j(h_j) \cap S_j(h'_j)$. This completes the proof of Claim 1.

Claim 2. Every two $h_j, h'_j \in H_j^{first}(C)$ are preceded by the same sequence of player j actions.

Proof of claim 2. If h_j and h'_j are not preceded by any player j actions, the statement is trivially true. Suppose now that h_j is preceded by a at least one player j action. Let $a_j^1, ..., a_j^K$ be the player j actions that precede h_j. We show that $a_j^1, ..., a_j^K$ also precede h'_j.

Suppose not. Then, there is some action $a_j^k \in A_j(h_j^k)$ that precedes h_j but not h'_j. We distinguish two cases: (1) h_j^k precedes h'_j, and (2) h_j^k does not precede h'_j.

Case 1. Suppose that h_j^k precedes h'_j. By Claim 1 there is some $s_j^* \in S_j(h_j) \cap S_j(h'_j)$. Since $s_j^* \in S_j(h_j)$ and a_j^k is the unique action at h_j^k that precedes h_j, we have that $s_j^*(h_j^k) = a_j^k$. Since $s_j^* \in S_j(h'_j)$ and h_j^k precedes h'_j it would follow that a_j^k precedes h'_j as well, which is a contradiction.

Case 2. Suppose that h_j^k does not precede h'_j. By Claim 1 there is some s_j^* in C with $s_j^* \in S_j(h_j) \cap S_j(h'_j)$. Take some $s_i \in S_i(h'_j)$. As $s_j^* \in S_j(h'_j)$ and, by perfect recall, $S(h'_j) = S_i(h'_j) \times S_j(h'_j)$, we conclude that $(s_i, s_j^*) \in S(h'_j)$. Since h_j^k does not precede h'_j it must be that $(s_i, s_j^*) \notin S(h_j^k)$. Recall that $a_j^k \in A_j(h_j^k)$ precedes h_j, which implies that h_j^k precedes h_j. Since $s_j^* \in S_j(h_j)$ it follows that $s_j^* \in S_j(h_j^k)$. As $(s_i, s_j^*) \notin S(h_j^k)$ and, by perfect recall, $S(h_j^k) = S_i(h_j^k) \times S_j(h_j^k)$, we conclude that $s_i \notin S_i(h_j^k)$.

Now, let t_j be a strategy that is minimally different from s_j^* at h_j^k. Since $s_i \notin S_i(h_j^k)$, it follows from Lemma 7.3 that $z(s_i, s_j^*) = z(s_i, t_j)$. Since $s_j^* \in C$ this would imply that $t_j \in C$ and $h_j^k \in H_j(C)$. However, this is a contradiction since h_j^k precedes $H_j^{first}(C)$, and can therefore not be in $H_j(C)$. We thus obtain a contradiction.

By Cases 1 and 2 we conclude that the actions $a_j^1, ..., a_j^K$ preceding h_j also precede h'_j. Hence, all player j actions that precede h_j also precede h'_j. In a similar fashion, it follows that all player j actions preceding h'_j also precede h_j. Thus, h_j and h'_j are preceded by the same player j actions. This completes the proof of Claim 2.

Claim 3. For every two $h_j, h'_j \in H_j^{first}(C)$ we have that $S_j(h_j) = S_j(h'_j)$.

Proof of Claim 3. By Claim 2, h_j and h'_j are preceded by the same player j actions $a_j^1, ..., a_j^K$ at the information sets $h_j^1, ..., h_j^K$. But then, by construction,

$$S_j(h_j) = \{s_j \in S_j \mid s_j(h_j^k) = a_j^k \text{ for all } k \in \{1, ..., K\}\} = S_j(h'_j).$$

This completes the proof of Claim 3.

We will now show that the induced graph G_i^D satisfies two strategies per connected component. Take a connected component C. We distinguish two cases: (1) $H_j^{first}(C)$ contains only one information set, and (2) $H_j^{first}(C)$ contains at least two information sets.

Case 1. Suppose that $H_j^{first}(C)$ contains a single information set h_j^*. As $h_j^* \in H_j(C)$ there are strategies s_j^*, t_j^* in C that are minimally different at h_j^* and a strategy s_i^* with $z(s_i^*, s_j^*) = z(s_i^*, t_j^*)$. By Lemma 7.3 we know that $s_i^* \notin S_i(h_j^*)$. As all other information sets in $H_j(C)$ follow h_j^* we conclude that $s_i^* \notin S_i(h_j)$ for every $h_j \in H_j(C)$.

Take an edge (s_j, t_j) in C. Hence, s_j and t_j are minimally different at some $h_j \in H_j(C)$ and there is some s_i with $z(s_i, s_j) = z(s_i, t_j)$. As we have seen above that $s_i^* \notin S_i(h_j)$, it follows by Lemma 7.3 that $z(s_i^*, s_j) = z(s_i^*, t_j)$. Thus, two strategies per connected component is satisfied. In fact, one strategy s_i^* turned out to be sufficient for the connected component C.

Case 2. Suppose that $H_j^{first}(C)$ contains at least two information sets h_j^1 and h_j^2. Choose a strategy $s_i^1 \in S_i(h_j^1)$ and a strategy $s_i^2 \in S_i(h_j^2)$.

Now, take an edge (s_j^*, t_j^*) in C. Then, s_j^*, t_j^* are minimally different at some $h_j^* \in H_j(C)$ and there is some strategy s_i with $z(s_i, s_j^*) = z(s_i, t_j^*)$. By definition of $H_j^{first}(C)$, information set h_j^* weakly follows some $h_j \in H_j^{first}(C)$. In fact, by perfect recall, h_j^* weakly follows exactly one information set in $H_j^{first}(C)$. We distinguish two cases: (2.1) h_j^* does not weakly follow h_j^1, and (2.2) h_j^* does not weakly follow h_j^2.

Case 2.1. Assume that h_j^* does not weakly follow h_j^1. Then, we show that $z(s_i^1, s_j^*) = z(s_i^1, t_j^*)$. Suppose that h_j^* weakly follows $h_j \in H_j^{first}(C) \backslash \{h_j^1\}$. As $s_j^* \in S_j(h_j^*)$ and h_j^* weakly follows h_j we conclude that $s_j^* \in S_j(h_j)$. Since we know, by Claim 3, that $S_j(h_j) = S_j(h_j^1)$ it follows that $s_j^* \in S_j(h_j^1)$. Recall from above that $s_i^1 \in S_i(h_j^1)$. Since, by perfect recall, $S(h_j^1) = S_i(h_j^1) \times S_j(h_j^1)$, we conclude that $(s_i^1, s_j^*) \in S(h_j^1)$. Since h_j^* does not weakly follow h_j^1 we conclude that $(s_i^1, s_j^*) \notin S(h_j^*)$. As s_j^*, t_j^* are minimally different at h_j^* it follows that $z(s_i^1, s_j^*) = z(s_i^1, t_j^*)$.

Case 2.2. Assume that h_j^* does not weakly follow h_j^2. Then, it can be shown in a similar fashion as above that $z(s_i^2, s_j^*) = z(s_i^2, t_j^*)$.

By Cases 2.1 and 2.2, the condition of two strategies per connected component is satisfied. Together with Case 1, we see that two strategies per connected component is satisfied whenever the game has two players and satisfies perfect recall. This completes the proof. ■

Lemma 7.5 (Strategy combinations leading to same consequence) *Consider a strategy s_i and two opponents' strategy combinations s_{-i}, t_{-i} with $z(s_i, s_{-i}) = z(s_i, t_{-i})$. Then, there are opponents' strategy combinations $s_{-i}^0, s_{-i}^1, ..., s_{-i}^K$ such that (i) $s_{-i}^0 = s_{-i}$, (ii) $s_{-i}^K = t_{-i}$, (iii) s_{-i}^k, s_{-i}^{k+1} minimally different for every $k \in \{0, ..., K-1\}$, and (iv) $z(s_i, s_{-i}^k) = z(s_i, s_{-i}^{k+1})$ for all $k \in \{0, ..., K-1\}$.*

Proof. Let the set of players be $I = \{1, ..., n\}$ and assume, without loss of generality, that $i = 1$. Let $s_{-i} = (s_2, ..., s_n)$ and $t_{-i} = (t_2, ..., t_n)$. For every opponent j let

$$H_j^{dif}(s_j, t_j) := \{h_j \in H_j(s_j) \cap H_j(t_j) \mid s_j(h_j) \neq t_j(h_j)\}$$

be the collection of information sets where s_j, t_j differ.

Take an opponent $j \in \{2, ..., n\}$, and suppose that $H_j^{dif}(s_j, t_j)$ consists of K_j information sets $\{h_j^1, ..., h_j^{K_j}\}$. We define strategies $s_j^0, ..., s_j^{K_j}$ as follows: Set $s_j^0 := s_j$, and for every $k \in \{1, ..., K_j\}$ let s_j^k be the unique strategy that (i) coincides with t_j at all information sets $h_j \in \{h_j^1, ..., h_j^k\}$, (ii) coincides with t_j at all information sets $h_j \in H_j(t_j)$ that follow an information set in $\{h_j^1, ..., h_j^k\}$, and (iii) coincides with s_j at all other information sets in $H_j(s_j^k)$. Then, by construction, $s_j^{K_j} = t_j$, and s_j^{k-1}, s_j^k are minimally different at h_j^k for every $k \in \{1, ..., K_j\}$.

For every $j \in \{2, ..., n\}$ and $k \in \{1, ..., K_j\}$ let

$$s_{-i}^{j.k} := (t_2, ..., t_{j-1}, s_j^k, s_{j+1}, ..., s_n).$$

Then, we define the sequence of opponents' strategy combinations $s_{-i}^0, s_{-i}^1, ..., s_{-i}^K$ by

$$s_{-i}^0, s_{-i}^1 ...., s_{-i}^K := s_{-i}, s_{-i}^{2.1}, ..., s_{-i}^{2.K_2}, s_{-i}^{3.1}, ..., s_{-i}^{3.K_3}, ..., s_{-i}^{n.1}, ..., s_{-i}^{n.K_n}.$$

By construction, $s_{-i}^0 = s_{-i}$, $s_{-i}^K = t_{-i}$ and s_{-i}^k, s_{-i}^{k+1} are minimally different for every $k \in \{0, ..., K-1\}$. It remains to show that $z(s_i, s_{-i}^k) = z(s_i, s_{-i}^{k+1})$ for every $k \in \{0, ..., K-1\}$.

Recall that $z(s_i, s_{-i}) = z(s_i, t_{-i})$. Let $z := z(s_i, s_{-i}) = z(s_i, t_{-i})$. Then, (s_i, s_{-i}) and (s_i, t_{-i}) select all the actions on the path to z. Now, take some $k \in \{0, ..., K-1\}$, and suppose that s_{-i}^k, s_{-i}^{k+1} minimally differ at some $h_j \in H_j$. Then, by construction, s_{-i}, t_{-i} also differ at h_j. Since (s_i, s_{-i}) and (s_i, t_{-i}) select all the actions on the path to z, it must be that h_j is not on the path to z. Hence, we conclude that $z(s_i, s_{-i}^k) = z(s_i, s_{-i}^{k+1}) = z$ also. Thus, $z(s_i, s_{-i}^k) = z(s_i, s_{-i}^{k+1})$ for every $k \in \{0, ..., K-1\}$. This completes the proof. ■

Lemma 7.6 (Induced probability distributions on consequences) *Consider two strategies $s_i, s_i' \in S_i$ and two beliefs β_i, β_i' with $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i')}$. Then, $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$.*

Proof. For every consequence z, let $S_i(z)$ be the set of strategies $s_i \in S_i$ that select all player i actions on the path to z, and let $S_{-i}(z)$ be the set of opponents' strategy combinations $s_{-i} \in S_{-i}$ that select all opponents' actions on the path to z. Take some consequence z with $\mathbb{P}_{(s_i,\beta_i)}(z) > 0$. Then, $s_i \in S_i(z)$ and $\mathbb{P}_{(s_i,\beta_i)}(z) = \beta_i(S_{-i}(z))$. As $\mathbb{P}_{(s_i',\beta_i')}(z) = \mathbb{P}_{(s_i,\beta_i)}(z) > 0$ we have that $s_i' \in S_i(z)$ and $\mathbb{P}_{(s_i',\beta_i')}(z) = \beta_i'(S_{-i}(z))$. Since $\mathbb{P}_{(s_i',\beta_i')}(z) = \mathbb{P}_{(s_i,\beta_i)}(z)$ it follows that $\beta_i(S_{-i}(z)) = \beta_i'(S_{-i}(z))$. But then, we conclude that

$$\mathbb{P}_{(s_i',\beta_i)}(z) = \beta_i(S_{-i}(z)) = \mathbb{P}_{(s_i,\beta_i)}(z) \text{ and } \mathbb{P}_{(s_i,\beta_i')}(z) = \beta_i'(S_{-i}(z)) = \beta_i(S_{-i}(z)) = \mathbb{P}_{(s_i,\beta_i)}(z).$$

As this holds for every z with $\mathbb{P}_{(s_i,\beta_i)}(z) > 0$, it follows that $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$. This completes the proof. ■

7.3 Utility Transformation Procedure

Consider a conditional preference relation $\succsim_i$ with an expected utility representation u_i. The following procedure, which we call the *utility transformation procedure,* transforms the utility function u_i into a new utility function v_i which is still an expected utility representation for $\succsim_i$ and that, under certain conditions, is a utility function on consequences. This procedure is used in the proofs of Theorems 4.1 and 5.1.

Take a dynamic game form D, a player i, and a conditional preference relation $\succsim_i$ for player i with an expected utility representation u_i. Recall from above the definition of the graph G_i^D induced by the dynamic game form D for player i, and fix a connected component C. Then, there is a spanning tree T for C with root s_{-i}^0 in C. If there are K nodes in C, choose a bijective numbering $m : C \to \{1, ..., K\}$ such that $m(s_{-i}) > m(t_{-i})$ whenever $s_{-i} \neq t_{-i}$ and t_{-i} lies on the unique path in T from s_{-i}^0 to s_{-i}. Hence, $m(s_{-i}^0) = 1$. We define the new utilities $v_i(s_i, s_{-i})$ for the nodes s_{-i} in C by induction on $m(s_{-i})$, as follows: For the node s_{-i}^0 with $m(s_{-i}^0) = 1$, set

$$v_i(s_i, s_{-i}^0) := u_i(s_i, s_{-i}^0) \tag{7.1}$$

for every strategy s_i.

Now, consider a node $s_{-i} \neq s_{-i}^0$ in C, and suppose that $v_i(s_i, t_{-i})$ has been defined for all strategies s_i and all nodes t_{-i} in C with $m(t_{-i}) < m(s_{-i})$. Consider the unique path in T from s_{-i}^0 to s_{-i}, and let $p(s_{-i})$ be the predecessor to s_{-i} on this path. Then, $m(p(s_{-i})) < m(s_{-i})$ which implies that $v_i(s_i, p(s_{-i}))$ has been defined for all strategies s_i. Moreover, let strategy $t_i(s_{-i})$ be such that $z(t_i(s_{-i}), p(s_{-i})) = z(t_i(s_{-i}), s_{-i})$. Define

$$v_i(s_i, s_{-i}) := u_i(s_i, s_{-i}) + v_i(t_i(s_{-i}), p(s_{-i})) - u_i(t_i(s_{-i}), s_{-i}) \tag{7.2}$$

for every strategy s_i. Then, by construction, $v_i(t_i(s_{-i}), s_{-i}) = v_i(t_i(s_{-i}), p(s_{-i}))$.

In this way we define the new utility $v_i(s_i, s_{-i})$ for every strategy s_i and every node s_{-i} in C. If we do so for every connected component C we define the new utility $v_i(s_i, s_{-i})$ for every strategy s_i and every opponents' strategy combination $s_{-i} \in S_{-i}$. The description of the new utility function v_i is hereby complete.

We will now show that the new utility function v_i still represents the conditional preference relation $\succsim_i$. On the basis of (7.1) and (7.2) we conclude that

$$v_i(s_i, s_{-i}) - v_i(t_i, s_{-i}) = u_i(s_i, s_{-i}) - u_i(t_i, s_{-i})$$

for every two strategies s_i, t_i and every node s_{-i}. As such, for every belief the expected utility difference between any two strategies will be the same under u_i as under v_i, which implies that v_i represents the same conditional preference relation as u_i. Since u_i represents the conditional preference relation $\succsim_i$, it follows that v_i represents $\succsim_i$ also.

7.4 Proof of Theorem 4.1

We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. (a) Suppose first that $\succsim_i$ is utility-based consequentialist. Then, $\succsim_i$ has an expected utility representation u_i on consequences. To show that $\succsim_i$ induces additive preference intensities on consequences, take two opponents' strategy combinations s^*_{-i}, t^*_{-i}, and two paths

$$s^*_{-i} \overset{s_i^1, z^1}{-} s^2_{-i} \overset{s_i^2, z^2}{-} s^3_{-i} ... \overset{s_i^{K-1}, z^{K-1}}{-} s^K_{-i} \overset{s_i^K, z^K}{-} t^*_{-i}$$

and

$$s^*_{-i} \overset{t_i^1, y^1}{-} t^2_{-i} \overset{t_i^2, y^2}{-} t^3_{-i} ... \overset{t_i^{L-1}, y^{L-1}}{-} t^L_{-i} \overset{t_i^L, y^L}{-} t^*_{-i}$$

from s^*_{-i} to t^*_{-i}. Then,

$$\begin{aligned} & \left[u_i(s_i^1, s^2_{-i}) - u_i(s_i^2, s^2_{-i})\right] + \left[u_i(s_i^2, s^3_{-i}) - u_i(s_i^3, s^3_{-i})\right] + ... \\ & ... + \left[u_i(s_i^{K-1}, s^K_{-i}) - u_i(s_i^K, s^K_{-i})\right] + \left[u_i(s_i^K, t^*_{-i}) - u_i(t_i^L, t^*_{-i})\right] \\ = \; & u_i(s_i^1, s^2_{-i}) - u_i(t_i^L, t^*_{-i}). \end{aligned} \tag{7.3}$$

Indeed, since $z(s_i^k, s^k_{-i}) = z(s_i^k, s^{k+1}_{-i})$ for all $k \in \{2, ..., K-1\}$ and $z(s_i^K, s^K_{-i}) = z(s_i^K, t^*_{-i})$, and u_i is a utility function on consequences, we have that $u_i(s_i^k, s^k_{-i}) = u_i(s_i^k, s^{k+1}_{-i})$ for all $k \in \{2, ..., K-1\}$ and $u_i(s_i^K, s^K_{-i}) = u_i(s_i^K, t^*_{-i})$.

In a similar fashion it follows that

$$\begin{aligned} & \left[u_i(s_i^1, s^*_{-i}) - u_i(t_i^1, s^*_{-i})\right] + \left[u_i(t_i^1, t^2_{-i}) - u_i(t_i^2, t^2_{-i})\right] + \\ & + \left[u_i(t_i^2, t^3_{-i}) - u_i(t_i^3, t^3_{-i})\right] + ... + \left[u_i(t_i^{L-1}, t^L_{-i}) - u_i(t_i^L, t^L_{-i})\right] \\ = \; & u_i(s_i^1, s^*_{-i}) - u_i(t_i^L, t^L_{-i}). \end{aligned} \tag{7.4}$$

Since $z(s_i^1, s^2_{-i}) = z(s_i^1, s^*_{-i})$ and $z(t_i^L, t^*_{-i}) = z(t_i^L, t^L_{-i})$, and u_i is a utility function on consequences, it follows that $u_i(s_i^1, s^2_{-i}) = u_i(s_i^1, s^*_{-i})$ and $u_i(t_i^L, t^*_{-i}) = u_i(t_i^L, t^L_{-i})$. If we combine this with (7.3) and (7.4) we conclude that

$$\begin{aligned} & \left[u_i(s_i^1, s^2_{-i}) - u_i(s_i^2, s^2_{-i})\right] + \left[u_i(s_i^2, s^3_{-i}) - u_i(s_i^3, s^3_{-i})\right] + ... \\ & ... + \left[u_i(s_i^{K-1}, s^K_{-i}) - u_i(s_i^K, s^K_{-i})\right] + \left[u_i(s_i^K, t^*_{-i}) - u_i(t_i^L, t^*_{-i})\right] \\ = \; & \left[u_i(s_i^1, s^*_{-i}) - u_i(t_i^1, s^*_{-i})\right] + \left[u_i(t_i^1, t^2_{-i}) - u_i(t_i^2, t^2_{-i})\right] + \\ & + \left[u_i(t_i^2, t^3_{-i}) - u_i(t_i^3, t^3_{-i})\right] + ... + \left[u_i(t_i^{L-1}, t^L_{-i}) - u_i(t_i^L, t^L_{-i})\right]. \end{aligned}$$

Hence, $\succsim_i$ induces additive preference intensities on consequences.

(b) Assume next that $\succsim_i$ has an expected utility representation u_i, induces additive preference intensities on consequences and has no weakly dominated strategies. Use the *utility transformation procedure* presented above to transform u_i into a new expected utility representation v_i. We will now show that v_i is a utility function on consequences.

Within the graph G_i^D, consider a connected component C and the associated spanning tree T with root s_{-i}^0 chosen in the utility transformation procedure. We prove, for every strategy s_i and every edge (s_{-i}^*, t_{-i}^*) in C that

$$v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*) \text{ whenever } z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*). \tag{7.5}$$

We distinguish two cases: (1) the edge (s_{-i}^*, t_{-i}^*) is in the spanning tree T, and (2) the edge (s_{-i}^*, t_{-i}^*) is not in the spanning tree T.

Case 1. Suppose that the edge (s_{-i}^*, t_{-i}^*) is in the spanning tree T with $t_{-i}^* = p(s_{-i}^*)$, where $p(s_{-i}^*)$ is the predecessor to s_{-i}^* in the utility transformation procedure. Take any strategy s_i with $z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*)$. Since $t_{-i}^* = p(s_{-i}^*)$ we know by (7.2) in the utility transformation procedure that $v_i(t_i(s_{-i}^*), s_{-i}^*) = v_i(t_i(s_{-i}^*), t_{-i}^*)$. As $z(t_i(s_{-i}^*), s_{-i}^*) = z(t_i(s_{-i}^*), t_{-i}^*)$ and $z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*)$, it follows by Lemma 7.2 that $v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*)$, and hence (7.5) holds.

Case 2. Suppose that the edge (s_{-i}^*, t_{-i}^*) is not in the spanning tree T. Let

$$s_{-i}^0 \overset{s_i^0, z^0}{-} s_{-i}^1 \overset{s_i^1, z^1}{-} s_{-i}^2 ... \overset{s_i^{L-1}, z^{L-1}}{-} s_{-i}^L \overset{s_i^L, z^L}{-} s_{-i}^* \tag{7.6}$$

be the unique path in T from s_{-i}^0 to s_{-i}^*. Moreover, let

$$s_{-i}^0 \overset{t_i^0, y^0}{-} t_{-i}^1 \overset{t_i^1, y^1}{-} t_{-i}^2 ... \overset{t_i^{M-1}, y^{M-1}}{-} t_{-i}^M \overset{t_i^M, y^M}{-} t_{-i}^*$$

be the unique path in T from s_{-i}^0 to t_{-i}^*.

As (s_{-i}^*, t_{-i}^*) is an edge, there is a strategy t_i such that $z(t_i, t_{-i}^*) = z(t_i, s_{-i}^*) =: y$. Then,

$$s_{-i}^0 \overset{t_i^0, y^0}{-} t_{-i}^1 \overset{t_i^1, y^1}{-} t_{-i}^2 ... \overset{t_i^{M-1}, y^{M-1}}{-} t_{-i}^M \overset{t_i^M, y^M}{-} t_{-i}^* \overset{t_i, y}{-} s_{-i}^* \tag{7.7}$$

is an alternative path from s_{-i}^0 to s_{-i}^*. Since $\succsim_i$ induces additive preference intensities on consequences, and v_i is an expected utility representation for $\succsim_i$, it follows from (7.6) and (7.7) that

$$\begin{aligned}
& \left[v_i(s_i^0, s_{-i}^1) - v_i(s_i^1, s_{-i}^1)\right] + \left[v_i(s_i^1, s_{-i}^2) - v_i(s_i^2, s_{-i}^2)\right] + ... \\
& ... + \left[v_i(s_i^{L-1}, s_{-i}^L) - v_i(s_i^L, s_{-i}^L)\right] + \left[v_i(s_i^L, s_{-i}^*) - v_i(t_i, s_{-i}^*)\right] \\
= \; & \left[v_i(s_i^0, s_{-i}^0) - v_i(t_i^0, s_{-i}^0)\right] + \left[v_i(t_i^0, t_{-i}^1) - v_i(t_i^1, t_{-i}^1)\right] + \\
& + \left[v_i(t_i^1, t_{-i}^2) - v_i(t_i^2, t_{-i}^2)\right] + ... + \left[v_i(t_i^{M-1}, t_{-i}^M) - v_i(t_i^M, t_{-i}^M)\right] + \left[v_i(t_i^M, t_{-i}^*) - v_i(t_i, t_{-i}^*)\right].
\end{aligned} \tag{7.8}$$

Note that all edges in (7.6) and (7.7), except $t^*_{-i} \overset{t_i,y}{-} s^*_{-i}$, are in T. By Case 1 it therefore follows that

$$\begin{gathered} v_i(s_i^k, s_{-i}^k) = v_i(s_i^k, s_{-i}^{k+1}) \text{ for all } k \in \{0, ..., L-1\},\ v_i(s_i^L, s_{-i}^L) = v_i(s_i^L, s^*_{-i}), \\ v_i(t_i^0, s_{-i}^0) = v_i(t_i^0, t_{-i}^1),\ v_i(t_i^k, t_{-i}^k) = v_i(t_i^k, t_{-i}^{k+1}) \text{ for all } k \in \{1, ..., M-1\} \text{ and} \\ v_i(t_i^M, t_{-i}^M) = v_i(t_i^M, t^*_{-i}). \end{gathered} \tag{7.9}$$

Combining (7.8) and (7.9) then yields $v_i(t_i, s^*_{-i}) = v_i(t_i, t^*_{-i})$.

Now, take an arbitrary s_i with $z(s_i, s^*_{-i}) = z(s_i, t^*_{-i})$. As we have seen above that $z(t_i, s^*_{-i}) = z(t_i, t^*_{-i})$ and $v_i(t_i, s^*_{-i}) = v_i(t_i, t^*_{-i})$, it follows by Lemma 7.2 that $v_i(s_i, s^*_{-i}) = v_i(s_i, t^*_{-i})$, and hence (7.5) holds. By Cases 1 and 2 we conclude that (7.5) holds for every edge (s^*_{-i}, t^*_{-i}) in the connected component C.

We finally show that the utility function v_i so constructed is a utility function on consequences. Take strategies s_i, t_i and opponents' strategy combinations s_{-i}, t_{-i} with $z(s_i, s_{-i}) = z(t_i, t_{-i})$. We will show that $v_i(s_i, s_{-i}) = v_i(t_i, t_{-i})$.

As $z(s_i, s_{-i}) = z(t_i, t_{-i}) =: z$, strategies s_i, t_i select all player i actions on the path to z, and s_{-i}, t_{-i} select all opponents' actions on the path to z. But then, $z(s_i, s_{-i}) = z(s_i, t_{-i})$ and $z(s_i, t_{-i}) = z(t_i, t_{-i})$.

As $z(s_i, s_{-i}) = z(s_i, t_{-i})$, it follows by Lemma 7.5 that we can choose opponents' strategy combinations $s^0_{-i}, s^1_{-i}, ..., s^M_{-i}$ such that (i) $s^0_{-i} = s_{-i}$, (ii) $s^M_{-i} = t_{-i}$, (iii) s^k_{-i}, s^{k+1}_{-i} are minimally different for every $k \in \{0, ..., M-1\}$, and (iv) $z(s_i, s^k_{-i}) = z(s_i, s^{k+1}_{-i})$ for all $k \in \{0, ..., M-1\}$. By (7.5) it then follows that $v_i(s_i, s^k_{-i}) = v_i(s_i, s^{k+1}_{-i})$ for all $k \in \{0, ..., M-1\}$, which implies that $v_i(s_i, s_{-i}) = v_i(s_i, t_{-i})$.

Moreover, as $z(s_i, t_{-i}) = z(t_i, t_{-i})$ it follows that $\mathbb{P}_{(s_i,[t_{-i}])} = \mathbb{P}_{(t_i,[t_{-i}])}$. Moreover, it trivially holds that $\mathbb{P}_{(s_i,[t_{-i}])} = \mathbb{P}_{(s_i,[t_{-i}])}$. Since $\succsim_i$ is preference-based consequentialist we know that

$$s_i \succsim_{i,[t_{-i}]} t_i \text{ if and only if } s_i \succsim_{i,[t_{-i}]} s_i.$$

Clearly, $s_i \sim_{i,[t_{-i}]} s_i$, and therefore $s_i \sim_{i,[t_{-i}]} t_i$. Since the utility function v_i represents $\succsim_i$ we must have that $v_i(s_i, t_{-i}) = v(t_i, t_{-i})$.

Together with the insight above that $v_i(s_i, s_{-i}) = v_i(s_i, t_{-i})$ we conclude that $v_i(s_i, s_{-i}) = v_i(t_i, t_{-i})$. As such, the utility function v_i is a utility function on consequences. Altogether, we have constructed an expected utility representation v_i for $\succsim_i$ that is a utility function on consequences. Hence, $\succsim_i$ is utility-based consequentialist. This completes the proof. ■

7.5 Proof of Theorem 5.1

Proof of Theorem 5.1. **(a)** Suppose first that $\succsim_i$ is utility-based consequentialist. Then, $\succsim_i$ has an expected utility representation u_i on consequences. Hence, for every consequence z there is a unique utility $\hat{u}_i(z)$ such that

$$u_i(s_i, s_{-i}) = \hat{u}_i(z) \text{ for all } (s_i, s_{-i}) \in S_i \times S_{-i} \text{ with } z(s_i, s_{-i}) = z.$$

For every strategy s_i and belief β_i we then have that

$$\begin{aligned} u_i(s_i, \beta_i) &= \sum_{s_{-i} \in S_{-i}} \beta_i(s_{-i}) \cdot u_i(s_i, s_{-i}) = \sum_{z \in Z} [\sum_{s_{-i} \in S_{-i} : z(s_i, s_{-i}) = z} \beta_i(s_{-i})] \cdot \hat{u}_i(z) \\ &= \sum_{z \in Z} \mathbb{P}_{(s_i, \beta_i)}(z) \cdot \hat{u}_i(z). \end{aligned} \tag{7.10}$$

To show that $\succsim_i$ is preference-based consequentialist, consider four strategies s_i, s_i', t_i, t_i' and two beliefs β_i, β_i' with

$$\mathbb{P}_{(s_i, \beta_i)} = \mathbb{P}_{(s_i', \beta_i')} \text{ and } \mathbb{P}_{(t_i, \beta_i)} = \mathbb{P}_{(t_i', \beta_i')}.$$

Then, in view of (7.10), $u_i(s_i, \beta_i) = u_i(s_i', \beta_i')$ and $u_i(t_i, \beta_i) = u_i(t_i', \beta_i')$, which implies that

$$u_i(s_i, \beta_i) - u_i(t_i, \beta_i) = u_i(s_i', \beta_i') - u_i(t_i', \beta_i').$$

Hence, $s_i \succsim_{i, \beta_i} t_i$ if and only if $s_i' \succsim_{i, \beta_i'} t_i'$. As such, $\succsim_i$ is preference-based consequentialist.

(b) Assume next that $\succsim_i$ has an expected utility representation u_i and is preference-based consequentialist. Use the *utility transformation procedure* to transform u_i into a new expected utility representation v_i for $\succsim_i$. We will now show that v_i is a utility function on consequences.

Within the graph G_i^D, consider a connected component C and the associated spanning tree T with root s_{-i}^0 chosen in the utility transformation procedure. We prove, for every strategy s_i and every edge (s_{-i}^*, t_{-i}^*) in C that

$$v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*) \text{ whenever } z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*). \tag{7.11}$$

We distinguish two cases: (1) the edge (s_{-i}^*, t_{-i}^*) is in the spanning tree T, and (2) the edge (s_{-i}^*, t_{-i}^*) is not in the spanning tree T.

Case 1. Suppose that the edge (s_{-i}^*, t_{-i}^*) is in the spanning tree T with $t_{-i}^* = p(s_{-i}^*)$, where $p(s_{-i}^*)$ is the predecessor to s_{-i}^* in the utility transformation procedure. Take any strategy s_i with $z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*)$. Then, it can be shown in the same way as in the proof of Theorem 4.1, Case 1, that $v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*)$, and hence (7.11) holds.

Case 2. Suppose that the edge (s_{-i}^*, t_{-i}^*) is not in the spanning tree T. Let $(s_{-i}^0, ..., s_{-i}^L)$ be the unique path in T from s_{-i}^0 to s_{-i}^*, where $s_{-i}^L = s_{-i}^*$. Moreover, let $(s_{-i}^{L+1}, ..., s_{-i}^{L+M})$ be the unique path in T from t_{-i}^* to s_{-i}^0, where $s_{-i}^{L+1} = t_{-i}^*$ and $s_{-i}^{L+M} = s_{-i}^0$. Then, $c := (s_{-i}^0, ..., s_{-i}^L, s_{-i}^{L+1}, ..., s_{-i}^{L+M})$ is a cycle in C.

By Lemma 7.4 we know that the graph G_i^D satisfies two strategies per connected component. Hence, there are two strategies s_i^*, t_i^* such that for every edge (s_{-i}^k, s_{-i}^{k+1}) in the cycle c either

$$z(s_i^*, s_{-i}^k) = z(s_i^*, s_{-i}^{k+1}) \text{ or } z(t_i^*, s_{-i}^k) = z(t_i^*, s_{-i}^{k+1}).$$

We distinguish three cases: (2.1) $z(s_i^*, s_{-i}^k) = z(s_i^*, s_{-i}^{k+1})$ for all edges (s_{-i}^k, s_{-i}^{k+1}) in the cycle c, (2.2) $z(t_i^*, s_{-i}^k) = z(t_i^*, s_{-i}^{k+1})$ for all edges (s_{-i}^k, s_{-i}^{k+1}) in the cycle c, and (2.3) conditions (2.1) and (2.2) do not hold.

Case 2.1. Suppose that $z(s_i^*, s_{-i}^k) = z(s_i^*, s_{-i}^{k+1})$ for all edges (s_{-i}^k, s_{-i}^{k+1}) in the cycle c. As the edges $(s_{-i}^0, s_{-i}^1), ..., (s_{-i}^{L-1}, s_{-i}^L)$ and the edges $(s_{-i}^{L+1}, s_{-i}^{L+2}), ..., (s_{-i}^{L+M-1}, s_{-i}^{L+M})$ are all in the spanning tree T, we know from Case 1 that

$$\begin{aligned} v_i(s_i^*, s_{-i}^*) &= v_i(s_i^*, s_{-i}^L) = v_i(s_i^*, s_{-i}^{L-1}) = ... = v_i(s_i^*, s_{-i}^0) \\ &= v_i(s_i^*, s_{-i}^{L+M}) = v_i(s_i^*, s_{-i}^{L+M-1}) = ... = v_i(s_i^*, s_{-i}^{L+1}) = v_i(s_i^*, t_{-i}^*). \end{aligned}$$

Hence, $v_i(s_i^*, s_{-i}^*) = v_i(s_i^*, t_{-i}^*)$.

Now, take an arbitrary s_i with $z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*)$. As $z(s_i^*, s_{-i}^*) = z(s_i^*, t_{-i}^*)$ and $v_i(s_i^*, s_{-i}^*) = v_i(s_i^*, t_{-i}^*)$, it follows by Lemma 7.2 that $v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*)$, and hence (7.11) holds.

Case 2.2. Suppose that $z(t_i^*, s_{-i}^k) = z(t_i^*, s_{-i}^{k+1})$ for all edges (s_{-i}^k, s_{-i}^{k+1}) in the cycle c. Then, it can be shown in the same way as in Case 2.1 that (7.11) holds for (s_{-i}^*, t_{-i}^*).

Case 2.3. Suppose that conditions (2.1) and (2.2) do not hold. Then, there is an edge (s_{-i}^k, s_{-i}^{k+1}) in the cycle c with $z(s_i^*, s_{-i}^k) \neq z(s_i^*, s_{-i}^{k+1})$ and an edge (s_{-i}^m, s_{-i}^{m+1}) with $z(t_i^*, s_{-i}^m) \neq z(t_i^*, s_{-i}^{m+1})$. Let

$$S_{-i}^+ := \{s_{-i}^k \text{ in } c \mid z(s_i^*, s_{-i}^{k-1}) \neq z(s_i^*, s_{-i}^k) \text{ and } z(s_i^*, s_{-i}^k) = z(s_i^*, s_{-i}^{k+1})\}$$

and

$$S_{-i}^- := \{s_{-i}^k \text{ in } c \mid z(s_i^*, s_{-i}^{k-1}) = z(s_i^*, s_{-i}^k) \text{ and } z(s_i^*, s_{-i}^k) \neq z(s_i^*, s_{-i}^{k+1})\},$$

where $s_{-i}^{-1} := s_{-i}^{M+L-1}$ and $s_{-i}^{M+L+1} := s_{-i}^1$. Then, S_{-i}^+ and S_{-i}^- are both non-empty, and have the same number of nodes, say n.

Define the beliefs

$$\beta_i^+ := \frac{1}{n} \sum_{s_{-i}^+ \in S_{-i}^+} [s_{-i}^+] \text{ and } \beta_i^- := \frac{1}{n} \sum_{s_{-i}^- \in S_{-i}^-} [s_{-i}^-].$$

Hence, β_i^+ assigns equal probability to all opponents' strategy combinations in S_{-i}^+, whereas β_i^- assigns equal probability to all opponents' strategy combinations in S_{-i}^-. We will show that

$$\mathbb{P}_{(s_i^*, \beta_i^+)} = \mathbb{P}_{(s_i^*, \beta_i^-)} \text{ and } \mathbb{P}_{(t_i^*, \beta_i^+)} = \mathbb{P}_{(t_i^*, \beta_i^-)}. \tag{7.12}$$

To prove this we introduce some additional notation. Fix the direction $(s_{-i}^0, ..., s_{-i}^L, s_{-i}^{L+1}, ..., s_{-i}^{L+M})$ of the cycle c. For every node $s_{-i}^+ \in S_{-i}^+$, let $fol(s_{-i}^+)$ be the first node in S_{-i}^- (given this direction) that follows s_{-i}^+, and let $pre(s_{-i}^+)$ be the last node in S_{-i}^- (given this direction) that precedes s_{-i}^+.

Now, consider some node $s_{-i}^+ \in S_{-i}^+$, and let $s_{-i}^k, s_{-i}^{k+1}, ..., s_{-i}^l$ be the sequence of nodes in c (if any) between s_{-i}^+ and $fol(s_{-i}^+)$ (given this direction). Then, by construction,

$$z(s_i^*, s_{-i}^+) = z(s_i^*, s_{-i}^k) = z(s_i^*, s_{-i}^{k+1}) = ... = z(s_i^*, s_{-i}^l) = z(s_i^*, fol(s_{-i}^+)). \tag{7.13}$$

Similarly, let $s_{-i}^m, s_{-i}^{m+1}, ..., s_{-i}^r$ be the sequence of nodes in c (if any) between $pre(s_{-i}^+)$ and s_{-i}^+ (given this direction). Then, by construction, $z(s_i^*, s_{-i}) \neq z(s_i^*, t_{-i})$ for every edge (s_{-i}, t_{-i}) on the path

$(pre(s^+_{-i}), s^m_{-i}, s^{m+1}_{-i}, ..., s^r_{-i}, s^+_{-i})$, and hence $z(t^*_i, s_{-i}) = z(t^*_i, t_{-i})$ for every edge (s_{-i}, t_{-i}) on the path $(pre(s^+_{-i}), s^m_{-i}, s^{m+1}_{-i}, ..., s^r_{-i}, s^+_{-i})$. As such,

$$z(t^*_i, pre(s^+_{-i})) = z(t^*_i, s^m_{-i}) = z(t^*_i, s^{m+1}_{-i}) = ... = z(t^*_i, s^r_{-i}) = z(t^*_i, s^+_{-i}). \tag{7.14}$$

We can then conclude that

$$\begin{aligned} \mathbb{P}_{(s^*_i, \beta^+_i)} &= \frac{1}{n} \sum_{s^+_{-i} \in S^+_{-i}} [z(s^*_i, s^+_{-i})] = \frac{1}{n} \sum_{s^+_{-i} \in S^+_{-i}} [z(s^*_i, fol(s^+_{-i}))] \\ &= \frac{1}{n} \sum_{s^-_{-i} \in S^-_{-i}} [z(s^*_i, s^-_{-i})] = \mathbb{P}_{(s^*_i, \beta^-_i)}. \end{aligned} \tag{7.15}$$

Here, the first equality follows from the definition of β^+_i, the second equality follows from (7.13), the third equality follows from the fact that

$$S^-_{-i} = \{fol(s^+_{-i}) \mid s^+_{-i} \in S^+_{-i}\},$$

whereas the last equality follows from the definition of β^-_i.

Similarly, it follows that

$$\begin{aligned} \mathbb{P}_{(t^*_i, \beta^+_i)} &= \frac{1}{n} \sum_{s^+_{-i} \in S^+_{-i}} [z(t^*_i, s^+_{-i})] = \frac{1}{n} \sum_{s^+_{-i} \in S^+_{-i}} [z(t^*_i, pre(s^+_{-i}))] \\ &= \frac{1}{n} \sum_{s^-_{-i} \in S^-_{-i}} [z(t^*_i, s^-_{-i})] = \mathbb{P}_{(t^*_i, \beta^-_i)}. \end{aligned} \tag{7.16}$$

Here, the first equality follows from the definition of β^+_i, the second equality follows from (7.14), the third equality follows from the fact that

$$S^-_{-i} = \{pre(s^+_{-i}) \mid s^+_{-i} \in S^+_{-i}\},$$

whereas the last equality follows from the definition of β^-_i. By (7.15) and (7.16) we thus conclude that (7.12) holds.

Since (i) (7.12) holds, (ii) the conditional preference relation $\succsim_i$ is preference-based consequentialist with expected utility representation v_i, and (iii) the two strategies s^*_i, t^*_i do not weakly dominate one another, we conclude on the basis of Lemma 7.1 that

$$v_i(s^*_i, \beta^+_i) - v_i(t^*_i, \beta^+_i) = v_i(s^*_i, \beta^-_i) - v_i(t^*_i, \beta^-_i). \tag{7.17}$$

By definition of the belief β^+_i we have that

$$v_i(s^*_i, \beta^+_i) = \frac{1}{n} \sum_{s^+_{-i} \in S^+_{-i}} v_i(s^*_i, s^+_{-i}),$$

and similarly for $v_i(t_i^*, \beta_i^+)$, $v_i(s_i^*, \beta_i^-)$ and $v_i(t_i^*, \beta_i^-)$. Substituting this into (7.17) yields

$$\frac{1}{n}\sum_{s_{-i}^+ \in S_{-i}^+} v_i(s_i^*, s_{-i}^+) - \frac{1}{n}\sum_{s_{-i}^+ \in S_{-i}^+} v_i(t_i^*, s_{-i}^+) = \frac{1}{n}\sum_{s_{-i}^- \in S_{-i}^-} v_i(s_i^*, s_{-i}^-) - \frac{1}{n}\sum_{s_{-i}^- \in S_{-i}^-} v_i(t_i^*, s_{-i}^-).$$

Since $S_{-i}^- = \{fol(s_{-i}^+) \mid s_{-i}^+ \in S_{-i}^+\}$ and $S_{-i}^- = \{pre(s_{-i}^+) \mid s_{-i}^+ \in S_{-i}^+\}$, this implies that

$$\sum_{s_{-i}^+ \in S_{-i}^+} v_i(s_i^*, s_{-i}^+) - \sum_{s_{-i}^+ \in S_{-i}^+} v_i(t_i^*, s_{-i}^+) = \sum_{s_{-i}^+ \in S_{-i}^+} v_i(s_i^*, fol(s_{-i}^+)) - \sum_{s_{-i}^+ \in S_{-i}^+} v_i(t_i^*, pre(s_{-i}^+)). \tag{7.18}$$

For every two nodes s_{-i}, t_{-i} on the cycle c, let $[s_{-i}, t_{-i}]$ be the ordered set of all the nodes on the cycle (including s_{-i} and t_{-i}) between s_{-i} and t_{-i} (in the direction of the cycle c). Recall the edge (s_{-i}^*, t_{-i}^*) on the cycle c we consider. Then, there is some node $s_{-i}^{*+} \in S_{-i}^+$ such that either $s_{-i}^*, t_{-i}^* \in [s_{-i}^{*+}, fol(s_{-i}^{*+})]$ or $s_{-i}^*, t_{-i}^* \in [pre(s_{-i}^{*+}), s_{-i}^{*+}]$. We thus distinguish two cases: (2.3.1) $s_{-i}^*, t_{-i}^* \in [s_{-i}^{*+}, fol(s_{-i}^{*+})]$ and (2.3.2) $s_{-i}^*, t_{-i}^* \in [pre(s_{-i}^{*+}), s_{-i}^{*+}]$.

Case 2.3.1. Assume that $s_{-i}^*, t_{-i}^* \in [s_{-i}^{*+}, fol(s_{-i}^{*+})]$. Take some $s_{-i}^+ \in S_{-i}^+ \backslash \{s_{-i}^{*+}\}$, and let

$$[s_{-i}^+, fol(s_{-i}^+)] = (s_{-i}^+, s_{-i}^1, ..., s_{-i}^k, fol(s_{-i}^+)).$$

Then, by construction,

$$z(s_i^*, s_{-i}^+) = z(s_i^*, s_{-i}^1) = ... = z(s_i^*, s_{-i}^k) = z(s_i^*, fol(s_{-i}^+)).$$

As all the edges in $[s_{-i}^+, fol(s_{-i}^+)]$ are in the spanning tree T, it follows by Case 1 that

$$v_i(s_i^*, s_{-i}^+) = v_i(s_i^*, s_{-i}^1) = ... = v_i(s_i^*, s_{-i}^k) = v_i(s_i^*, fol(s_{-i}^+)) \tag{7.19}$$

for all $s_{-i}^+ \in S_{-i}^+ \backslash \{s_{-i}^{*+}\}$.

Next, take some $s_{-i}^+ \in S_{-i}^+$, possibly equal to s_{-i}^{*+}, and let

$$[pre(s_{-i}^+), s_{-i}^+] = (pre(s_{-i}^+), s_{-i}^1, ..., s_{-i}^l, s_{-i}^+).$$

Then, by construction,

$$z(t_i^*, pre(s_{-i}^+)) = z(t_i^*, s_{-i}^1) = ... = z(t_i^*, s_{-i}^l) = z(t_i^*, s_{-i}^+).$$

As all the edges in $[pre(s_{-i}^+), s_{-i}^+]$ are in the spanning tree T, it follows by Case 1 that

$$v_i(t_i^*, pre(s_{-i}^+)) = v_i(t_i^*, s_{-i}^1) = ... = v_i(t_i^*, s_{-i}^l) = v_i(t_i^*, s_{-i}^+) \tag{7.20}$$

for all $s_{-i}^+ \in S_{-i}^+$.

By (7.19) and (7.20) we then conclude that all terms in (7.18) cancel, except for $v_i(s_i^*, s_{-i}^{*+})$ and $v_i(s_i^*, fol(s_{-i}^{*+}))$, which yields

$$v_i(s_i^*, s_{-i}^{*+}) = v_i(s_i^*, fol(s_{-i}^{*+})). \tag{7.21}$$

Recall that $s_{-i}^*, t_{-i}^* \in [s_{-i}^{*+}, fol(s_{-i}^{*+})]$ where t_{-i}^* follows s_{-i}^* in the direction of the cycle. Then, every edge (s_{-i}, t_{-i}) in $[s_{-i}^{*+}, s_{-i}^*]$, if any, is in the spanning tree T. As $z(s_i^*, s_{-i}) = z(s_i^*, t_{-i})$ for every such edge, it follows from Case 1 that $v_i(s_i^*, s_{-i}) = v_i(s_i^*, t_{-i})$ for every edge (s_{-i}, t_{-i}) in $[s_{-i}^{*+}, s_{-i}^*]$, if any. As such,

$$v_i(s_i^*, s_{-i}^*) = v_i(s_i^*, s_{-i}^{+*}). \tag{7.22}$$

Similarly, every edge (s_{-i}, t_{-i}) in $[t_{-i}^*, fol(s_{-i}^{*+})]$, if any, is in the spanning tree T. As $z(s_i^*, s_{-i}) = z(s_i^*, t_{-i})$ for every such edge, it follows from Case 1 that $v_i(s_i^*, s_{-i}) = v_i(s_i^*, t_{-i})$ for every edge (s_{-i}, t_{-i}) in $[t_{-i}^*, fol(s_{-i}^{*+})]$, if any. As such,

$$v_i(s_i^*, t_{-i}^*) = v_i(s_i^*, fol(s_{-i}^{*+})). \tag{7.23}$$

By (7.21), (7.22) and (7.23) it follows that $v_i(s_i^*, s_{-i}^*) = v_i(s_i^*, t_{-i}^*)$.

Now, take some arbitrary s_i with $z(s_i, s_{-i}^*) = z(s_i, t_{-i}^*)$. As $z(s_i^*, s_{-i}^*) = z(s_i^*, t_{-i}^*)$ and $v_i(s_i^*, s_{-i}^*) = v_i(s_i^*, t_{-i}^*)$, it follows from Lemma 7.2 that $v_i(s_i, s_{-i}^*) = v_i(s_i, t_{-i}^*)$. Hence, (7.11) holds.

Case 2.3.2. Assume that $s_{-i}^*, t_{-i}^* \in [pre(s_{-i}^{*+}), s_{-i}^{*+}]$. Then, it can be shown in a similar fashion as in Case 2.3.1 that (7.11) holds.

As we have exhausted all cases, we conclude that (7.11) holds for every edge (s_{-i}^*, t_{-i}^*) in the connected component C. Moreover, by covering all connected components C, we conclude that (7.11) holds for every edge (s_{-i}^*, t_{-i}^*) in the graph G_i^D.

In the same way as in the proof of Theorem 4.1 it can then be shown that the utility function v_i so constructed is a utility function on consequences. Hence, $\succsim_i$ is utility-based consequentialist. This completes the proof. ■

7.6 Proof of Theorem 4.2

Proof of Theorem 4.2. (a) Suppose first that $\succsim_i$ is preference-based consequentialist. Take two strategies s_i, t_i, and consider the restricted conditional preference relation $\succsim_i^{\{s_i,t_i\}}$. Then, $\succsim_i^{\{s_i,t_i\}}$ is preference-based consequentialist also. As $\succsim_i^{\{s_i,t_i\}}$ only involves two strategies, it follows from the proof of Theorem 5.1, part (b), that $\succsim_i^{\{s_i,t_i\}}$ is utility-based consequentialist. By Theorem 4.1 we then conclude that $\succsim_i^{\{s_i,t_i\}}$ induces additive preference intensities on consequences.

(b) Suppose next that for every pair of strategies s_i, t_i, the restricted conditional preference relation $\succsim_i^{\{s_i,t_i\}}$ induces additive preference intensities on consequences. Let u_i be an expected utility representation of $\succsim_i$. Take a pair of strategies s_i, t_i. As $\succsim_i^{\{s_i,t_i\}}$ induces additive preference intensities on consequences, it follows from Theorem 4.1 that $\succsim_i^{\{s_i,t_i\}}$ is utility-based consequentialist. By the proof of Theorem 5.1, part (a), it follows that $\succsim_i^{\{s_i,t_i\}}$ is preference-based consequentialist. Thus, $\succsim_i^{\{s_i,t_i\}}$ is preference-based consequentialist for every two strategies s_i, t_i.

We will now show that $\succsim_i$ is preference-based consequentialist. Take four strategies s_i, s_i', t_i, t_i' and two beliefs β_i, β_i' with $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i')}$ and $\mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i',\beta_i')}$. Then, it follows from Lemma 7.6 that $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i,\beta_i')}$ and $\mathbb{P}_{(t_i,\beta_i)} = \mathbb{P}_{(t_i,\beta_i')}$. Since $\succsim_i^{\{s_i,t_i\}}$ is preference-based consequentialist, it follows from Lemma 7.1 that

$$u_i(s_i,\beta_i) - u(t_i,\beta_i) = u_i(s_i,\beta_i') - u_i(t_i,\beta_i'). \tag{7.24}$$

Moreover, as $\mathbb{P}_{(s_i,\beta_i)} = \mathbb{P}_{(s_i',\beta_i')}$ we know from Lemma 7.6 that $\mathbb{P}_{(s_i,\beta_i')} = \mathbb{P}_{(s_i',\beta_i')}$. As, trivially, $\mathbb{P}_{(s_i,\beta_i')} = \mathbb{P}_{(s_i,\beta_i')}$, and $\succsim_i^{\{s_i,s_i'\}}$ is preference-based consequentialist, it follows from Lemma 7.1 that

$$u_i(s_i,\beta_i') - u_i(s_i',\beta_i') = u_i(s_i,\beta_i') - u_i(s_i,\beta_i') = 0,$$

which implies that $u_i(s_i,\beta_i') = u_i(s_i',\beta_i')$. Similarly, it can be shown that $u_i(t_i,\beta_i') = u_i(t_i',\beta_i')$. Combining the latter two insights with (7.24) yields $u_i(s_i,\beta_i) - u_i(t_i,\beta_i) = u_i(s_i',\beta_i') - u_i(t_i',\beta_i')$. Hence, $s_i \succsim_{i,\beta_i} t_i$ if and only if $s_i' \succsim_{i,\beta_i'} t_i'$. We thus conclude that $\succsim_i$ is preference-based consequentialist. This completes the proof. ■

References

[1] Anscombe, F.J. and R.J. Aumann (1963), A definition of subjective probability, *Annals of Mathematical Statistics* **34,** 199–205.

[2] Gilboa, I. and D. Schmeidler (2003), A derivation of expected utility maximization in the context of a game, *Games and Economic Behavior* **44,** 184–194.

[3] Grant. S. (1995), Subjective probability without monotonicity: Or how Machina's mom may also be probabilistically sophisticated, *Econometrica* **63,** 159–189.

[4] Kuhn, H. (1953), Extensive games and the problem of information, in H.W. Kuhn and A.W. Tucker (eds.), *Contributions to the Theory of Games,* Volume II (Princeton University Press, Princeton, NJ, 1953), pp. 193–216 (*Annals of Mathematics Studies* **28**).

[5] Machina, M.J. (1989), Dynamic consistency and non-expected utility models of choice under uncertainty, *Journal of Economic Literature* **XXVII,** 1622–1668.

[6] Machina, M.J. and D. Schmeidler (1992), A more robust definition of subjective probability, *Econometrica* **60,** 745–780.

[7] Osborne, M.J. and A. Rubinstein (1994), *A Course in Game Theory,* The MIT Press, Cambridge, Massachusetts, London, England.

[8] Perea, A. (2023), Expected utility as an expression of linear preference intensity, Manuscript downloadable at: `www.epicenter.name/Perea/Papers/EU-linear-pref-int.pdf`

[9] Rubinstein, A. (1991), Comments on the interpretation of game theory, *Econometrica* **59**, 909–924.

[10] Savage, L.J. (1954), *The Foundation of Statistics,* Wiley, New York.

[11] Sinnott-Armstrong, W. (2023), Consequentialism, *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), `https://plato.stanford.edu/archives/win2023/entries/consequentialism/`

[12] von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.