



WINPEC Working Paper Series No.E2208

January 2023

# Behavioral bargaining theory: Equality bias, risk attitude, and reference-dependent utility

Yoshio Kamijo     Koji Yokote

Waseda INstitute of Political EConomy  
Waseda University  
Tokyo, Japan

# Behavioral bargaining theory: Equality bias, risk attitude, and reference-dependent utility

Yoshio Kamijo\*

Koji Yokote†

2023-01-10

## Abstract

We develop a new theory, termed the *behavioral bargaining theory* (henceforth, BBT), that explains various observed behaviors in bargaining experiments in a unified manner. The key idea is to modify Nash's (1950) model by endowing the players' utility functions with a new concept, named *entitlement*, that represents the amount of money the player feels entitled to receive. We first apply BBT to explain the *equality bias* that is widely observed in the laboratory. We argue that our explanation of the bias in terms of entitlements is more easily interpretable than the extant explanation in terms of risk attitudes. Then, we demonstrate that BBT can also explain other behavioral patterns beyond the equality bias by suitably setting entitlements. Finally, we provide empirical support to BBT by using experimental data from Takeuchi et al. (2022), where entitlements of players can be inferred from the experimental design.

**Keywords:** Behavioral bargaining theory, Nash bargaining solution, Reference dependent utility, Equality bias, Equal-split norm

---

\*Waseda University, yoshio.kamijo@gmail.com

†JSPS Research Fellow, Graduate School of Economics, the University of Tokyo, koji.yokote@gmail.com

## 1. Introduction

Bargaining is one of the most pervasive forms of economic transactions. It is often conducted between two parties with opposing interests, e.g., employers and employees, buyers and sellers, or developers and landowners, in pursuit of an agreed-upon outcome. In his seminal paper in 1950, Nash developed a model for analyzing bargaining problems in a unified manner and introduced the *Nash solution* (Nash, 1950). His model has become a cornerstone for the analysis of bargaining both in theory and in applications.

Although Nash’s model stands out in its generality and elegance of axiomatic approach, its relevance to real-world problems has long been questioned.<sup>1</sup> Skepticism has even been hardened as experimental data deviating from the Nash solution’s predictions has accumulated. To overcome this limitation, this paper develops a new bargaining theory, termed the *behavioral bargaining theory* (henceforth, BBT), that can explain the subjects’ behavioral patterns in a unified manner.

Critics often argue that the Nash solution fails to explain the players’ behaviors because it does not consider psychological factors, such as fairness or an “aspiration level” (Luce and Raiffa, 1989). Inspired by this criticism, we modify Nash’s model by endowing utility functions with a new term that captures psychological factors, which we call *entitlement*. Conceptually, our formulation of utility functions is similar to that of *reference-based preferences* advanced by Tversky and Kahneman (1979) and Köszegi and Rabin (2006). While the reference point is typically interpreted as some standard against which things are compared, we grant it a more concrete interpretation: it is interpreted as the amount of money that the player feels entitled to receive. The key step of BBT is to calculate the Nash solution under entitlement-dependent utility functions while inferring entitlements from past data or experimental design.

To demonstrate how the new theory explains actual behaviors, we first focus on the most well-known deviation of the Nash solution from reality, namely, *the equality bias*: subjects tend to choose a bargaining outcome away from the Nash solution toward 50–50 sharing (Anbarci and Feltovich, 2013, 2018; Birkeland and Tungodden, 2014; Hoffman and Spitzer, 1982; Nydegger and Owen, 1974; Roth, 1995). Within the framework of Nash’s original model, this bias is explained in terms of risk attitudes. Existing studies have revealed that if players are risk-loving, then the Nash solution shifts toward equal sharing.<sup>2</sup> We strengthen this claim to the following theorem, termed the *equality bias theorem*: if the players’ utility functions exhibit *increasing absolute risk aversion* (shortly, IARA), then the equality bias occurs; if utilities exhibit *decreasing absolute risk aversion* (shortly, DARA), then the *inequality bias* occurs. Therefore, the key driving force behind the shift toward an equal split of a pie is not the players’ risk attitudes but rather how risk attitudes change in response to the amount of money.<sup>3</sup>

BBT explains the equality bias differently from Nash’s original model by incorporating psychological factors, most notably the 50–50 norm or the equal split norm. This norm induces players to regard equitable sharing as a reference point for bargaining, and its existence has been verified in various contexts (Andreoni and Bernheim, 2009). Setting the players’ entitlements to be the equal split of the total pie and assuming linear utilities, we derive the allocation of the Nash solution and call it the *egalitarian neutral Nash allocation* (abbreviated as ENNA). Our second theorem states that equal sharing arises as a result of this solution. We further demonstrate that ENNA is consistent with bargaining outcomes in past experimental data (Anbarci and Feltovich, 2018). Comparing the explanation of the bias from Nash’ original model with that from BBT,

---

<sup>1</sup>For example, in the classical textbook of game theory by Luce and Raiffa (1989), they devote one chapter (Section 6.6) to criticizing the Nash solution.

<sup>2</sup>See, for example, Exercise 15.21 of Maschler et al. (2013).

<sup>3</sup>We remark that there exists a concave and IARA utility function, which exhibits risk-aversion but the bargaining outcome shifts toward the equal split.

we argue that the latter offers a more easily interpretable result.

Next, we turn our attention to other behavioral patterns beyond the equality bias. BBT is flexible enough to explain existing experimental data in a unified manner. Specifically, we take up three experiments. The first experiment is when players have no entitlements in mind, namely, when the bargaining outcome is designed to be a “windfall gain”; subjects are just offered a bargaining pie and notified of a disagreement outcome without being exposed to any other information. We show that when utility functions exhibit IARA (resp. DARA), the Nash solution exhibits the equality (resp. inequality) bias. This result suggests that, under IARA utilities, our model is consistent with existing experimental findings. The second experiment is the opposite case, namely when the entitlements coincide with the disagreement payoffs. When subjects need to earn the disagreement payoffs on their own through a costly task, they are induced to regard the payoff as their entitlement. We prove that, regardless of the assumption on risk attitudes, the Nash solution coincides with the equal split of the surplus (i.e., the total bargaining pie minus the disagreement payoffs). This result also finds empirical support. The third experiment is bargaining over a loss. Bargaining is often conducted not over a profit but over a loss, as in the case of the seminal *bankruptcy problem* (O’Neill, 1982; Aumann and Maschler, 1985). In our model with entitlements, a loss can be represented as receiving less money than the subject’s entitlement. It turns out that the Nash solution’s outcome under IARA utilities is consistent with the dictation of existing normative allocation rules. In contrast, the outcome under DARA utilities is rather counter-intuitive: the player with fewer entitlements may receive a larger pie, which we call the *entitlement paradox*. We draw insights from this theoretical result for experimental outcomes in the literature.

The important feature of entitlements is that they can be inferred from past data or experimental design. This contrasts sharply with other parameters, such as the disagreement points or utility functions, that are typically unobservable. To exploit this advantage, we apply BBT to a previous experiment where the subjects’ entitlements can be inferred and statistically test its explanatory power. We borrow experimental data from Takeuchi et al. (2022), who conducted experiments of bargaining in which the subjects first engage with costly tasks, and then the total bargaining pie is determined based on their effort levels. We infer the subjects’ entitlements from two different sources. The first source is observed data. We perform maximum likelihood estimation to identify the parameters of entitlements such that the theoretical prediction of BBT achieves the best data fitting. The second source is post-experiment questionnaires that elicit the subjects’ entitlements. We observe that the entitlements estimated from these two different sources exhibit a remarkable similarity. The analysis also highlights the advantage of BBT: by extracting essential components in bargaining situations (i.e., players’ entitlements), the theory offers a versatile and easy-to-use tool for interpreting bargaining behaviors.

The paper is organized as follows. In the next section, we briefly explain the setting of a simple bargaining problem. The explanation of the equality bias from the traditional approach is in Section 3. In Section 4, we introduce reference-dependent utility and entitlements, the fundamental concepts of our new bargaining theory. As the first application of BBT, we discuss an equal split norm and bargaining outcomes affected by the norm in Section 5. In Section 6, we explain other applications of BBT, which include the bargaining for the manna from heaven, the bargaining based on the earned disagreement payoff, and the bargaining for a loss. Section 7 deals with data-fitting, and Section 8 concludes the paper.

## 2. Model

A *bargaining problem* consists of players, a potential profit of agreement, and a disagreement outcome. There are two players 1 and 2. They bargain over a fixed amount  $M$  of some divisible good in pursuit of an agreed-upon outcome  $(x_1, x_2)$  with  $x_1 + x_2 = M$ . If the bargaining breaks

down, each gets  $v_1 \geq 0, v_2 \geq 0$  of the divisible good. A disagreement outcome is  $v = (v_1, v_2)$ . Let  $u_i$  ( $i = 1, 2$ ) denote  $i$ 's utility function and  $d = (d_1, d_2) = (u_1(v_1), u_2(v_2))$  denote the utility profile at the disagreement outcome. The set of possible pairs of utilities through bargaining is given by

$$Z = \{(z_1, z_2) : \exists(x_1, x_2) \text{ with } x_1 + x_2 = M \text{ such that} \\ (z_1, z_2) = (u_1(x_1), u_2(x_2)) \text{ and } (z_1, z_2) \geq d\}.$$

We denote a bargaining problem by  $(M, v)$  without referring to players, which are clear from the context.

A bargaining solution is a function that chooses an element in  $Z$  for any bargaining problem. One of the most eminent solutions is the *Nash solution*, which has several normative properties and positive interpretations (Binmore et al., 1986; Young, 1993; Rubinstein et al., 1992). The Nash solution chooses a utility pair that maximizes the product of the players' utility differences between an agreed-upon outcome and the disagreement outcome (this product is called the *Nash product*). Formally, the Nash solution chooses a solution to the following problem:

$$\begin{aligned} \max \quad & (u_1(x_1) - u_1(v_1)) \times (u_2(x_2) - u_2(v_2)) \\ \text{s.t.} \quad & x_1 + x_2 = M, \quad x_1 \geq v_1, \quad x_2 \geq v_2. \end{aligned}$$

Letting  $(x_1^*, x_2^*)$  denote a solution to the above problem,<sup>4</sup> the Nash solution is written as

$$NS(Z, d) = (u_1(x_1^*), u_2(x_2^*)).$$

Although this solution is defined on a utility basis, we are more interested in the allocation  $(x_1^*, x_2^*)$ . We refer to the pair as the *Nash allocation* and write

$$NA(M, v) = (x_1^*, x_2^*).$$

The *neutral Nash solution* is the Nash solution when the players' utility functions are risk-neutral. The allocation under the neutral Nash solution, called the *neutral Nash allocation*, is defined by

$$NNA(M, v) = \left( v_1 + \frac{M - v_1 - v_2}{2}, v_2 + \frac{M - v_1 - v_2}{2} \right).$$

The neutral Nash solution is often used in the literature in order to abstract away the effect of utility functions on the bargaining outcome. This solution is also known as the *equal difference solution* in bargaining theory and the *equal surplus solution* (or the *standard solution*) in cooperative game theory.

An important reference point of bargaining outcomes is the *equal split allocation* defined by

$$EA(M, v) = (M/2, M/2).$$

We briefly explain why we focus on the above three allocations (NA, NNA, and EA). Our goal is to examine whether observed bargaining outcomes can be explained as the outcome of the Nash solution. As is the case in experiments, we assume that  $v_i$  ( $i = 1, 2$ ) and  $M$  are observable, but  $u_i$  ( $i = 1, 2$ ) are unobservable; thus, NA cannot be directly computed. To overcome this difficulty, we instead compute NNA and EA (which can be done only by using observable information), and then identify the positional relationship between these allocations and NA. It will turn out

---

<sup>4</sup>We later impose assumptions that guarantee the uniqueness of  $(x_1^*, x_2^*)$

that there is indeed a clear-cut relationship if utility functions satisfy certain assumptions.

Suppose that  $u_i$  is defined over some large interval  $[x_{min}, x_{max}]$ ; usually we set  $x_{min} = 0$  and  $x_{max} = \infty$ . For  $i = 1, 2$ ,  $u_i$  is assumed to satisfy the following:

**A1.**  $u_i(0) = 0$  and  $u_i(x_i) > 0$ ,  $u'_i(x_i) > 0 \ \forall x_i > 0$ .

**A2.** (Increasing fear of ruin)  $(u'_i(x_i))^2 > u_i(x_i)u''_i(x_i) \ \forall x_i > 0$ .

The second is weaker than risk aversion because  $u''_i(x_i) < 0$  and A1 imply A2.<sup>5</sup> This generalization is important when we allow for risk-loving agents.

We explain the intended meaning of A2.<sup>6</sup> Aumann and Kurz (1977) refer to  $u(x)/u'(x)$  as *fear of ruin*. Consider a person who is going to bet all his money  $x$  against a small profit  $h$  with probability  $1 - q$ . Let  $q(h)$  denote the value of  $q$  with which she is indifferent between taking this gamble and keeping her current wealth. Then, the smaller the  $q(h)$  is, the more she fears ruin. Extending this argument to a local behavior of the probability per dollar, we define fear of ruin as the inverse of  $\lim_{h \rightarrow 0} q(h)/h$ .

Formally,<sup>7</sup>  $u(x) = (1 - q(h))u(x + h) + q(h)u(0)$  implies  $q(h) = (u(x + h) - u(x))/u(x + h)$ , which in turn implies

$$\lim_{h \rightarrow 0} \frac{q(h)}{h} = \lim_{h \rightarrow 0} \frac{(u(x + h) - u(x))/h}{u(x + h)} = \frac{u'(x)}{u(x)}$$

The inverse of this value,  $u(x)/u'(x)$ , is her fear of ruin. Notice that A2 requires this value to be increasing because

$$(u(x)/u'(x))' = \frac{(u'(x))^2 - u(x)u''(x)}{(u'(x))^2} > 0 \iff (u'(x))^2 - u(x)u''(x) > 0.$$

Increasing fear of ruin is also equivalent to log-concavity of  $u(x)$  because

$$\frac{d^2 \log(u(x))}{dx^2} = \frac{d}{dx} \left( \frac{u'(x)}{u(x)} \right) = \frac{u''(x)u(x) - (u'(x))^2}{(u(x))^2}.$$

The *Arrow-Pratt coefficient* at wealth level  $x$  is defined by

$$r(x) = -\frac{u''(x)}{u'(x)} = -\frac{d}{dx} \log u'(x).$$

This measure is closely related to the person's local risk attitude at  $x$ . The larger  $r(x)$  is, the larger its certainty equivalent and willingness to pay for an insurance (Pratt (1964), Theorem 1 p 128). In addition, any utility function can be derived from its Arrow-Pratt coefficient  $r(\cdot)$  as follows: by integrating  $r(\cdot)$  we have  $\log u'(x)$ , and then by taking the exponential of this and integrating again we have a positive affine transformation of  $u(x)$ .

Finally, we define a global property of risk attitude. A utility function  $u$  exhibits:

- *increasing absolute risk averse* (IARA) if  $r'(x) > 0$  for all  $x$  (i.e., the Arrow-Pratt coefficient is increasing as wealth increases).
- *decreasing absolute risk averse* (DARA) if  $r'(x) < 0$  for all  $x$ .
- *constant absolute risk aversion* (CARA) if  $r'(x) = 0$  for all  $x$ .

<sup>5</sup>Svejnar (1986) showed that risk aversion is a sufficient condition for the increasing fear of ruin.

<sup>6</sup>For simplicity, we write  $u(x)$  rather than  $u_i(x_i)$ .

<sup>7</sup>Here, we use A1 to guarantee  $u(x) \neq 0$  and  $u'(x) \neq 0$ .

It is well known that  $u(x) = x^\alpha$  is IARA if  $\alpha > 1$ , CARA if  $\alpha = 1$ , and DARA if  $0 < \alpha < 1$ . Using the terminology of log-concavity and log-convexity, IARA (resp. DARA) is equivalent to log-concavity (resp. log-convexity) of  $u'(x)$  because  $r'(x) = -\frac{d^2}{dx^2} \log u'(x)$ .

### 3. Explanation of the equality bias from canonical utility functions

As noted in Section 2, our goal is to identify the positional relationship between the Nash allocation and other benchmark allocations. To this end, we conduct a comparative statics analysis of the Nash solution when the total amount  $M$  changes.

First, we briefly summarize the basic properties of the Nash allocation  $(x_1^*, x_2^*)$ .

**P1.** When  $M = v_1 + v_2$ ,  $x_1^* = v_1, x_2^* = v_2$ .

**P2.** When  $M > v_1 + v_2$ ,  $x_1^* > v_1, x_2^* > v_2$ .

These two properties are obvious from the definition of the Nash solution being a maximizer of the Nash product. An important implication of P2 is that the Nash allocation is always obtained as an interior solution. Thus, FOC of the Nash product induces the following property:

**P3.** It holds that

$$u_1'(x_1^*)[u_2(x_2^*) - u_2(v_2)] = u_2'(x_2^*)[u_1(x_1^*) - u_1(v_1)].$$

With these basic properties at hand, we prove the following three technical properties (their proofs are in the Appendix).

**P4.** Let  $A = [u_1(x_1^*) - u_1(v_1)], B = [u_2(x_2^*) - u_2(v_2)]$ . If  $M > 0$ , we have

$$u_1'(x_1^*)u_2'(x_2^*) - u_2''(x_2^*)A > 0, \quad u_1'(x_1^*)u_2'(x_2^*) - u_1''(x_1^*)B > 0$$

**P5.** For any  $(M, v)$  with  $M \geq v_1 + v_2$ , under the assumptions of A1 and A2, the Nash allocation  $(x_1^*, x_2^*)$  is uniquely determined. In addition,  $x_i^*$  is continuous in  $v_1, v_2$  and  $M$ .

The following property is related to a locus of the bargaining allocation  $(x_1^*, x_2^*)$  when  $M$  increases with keeping  $v$  fixed.

**P6.** The locus of  $(x_1^*, x_2^*)$  when  $M$  increases with keeping  $v$  fixed is an ascending right curve starting at  $(v_1, v_2)$ , and the slope of the locus is give by

$$\frac{dx_2^*}{dx_1^*} = \frac{u_1(x_1^*)u_2(x_2^*) - u_1''(x_1^*)B}{u_1(x_1^*)u_2(x_2^*) - u_2''(x_2^*)A}, \quad (1)$$

where  $A = u_1(x_1^*) - u_1(v_1)$  and  $B = u_2(x_2^*) - u_2(v_2)$ .

We are now ready to present an important lemma on the comparative statics of the Nash allocation; more specifically, it clarifies how the Nash allocation  $(x_1^*, x_2^*)$  changes as the size of the bargaining pie  $M$  changes while fixing  $(v_1, v_2)$ .

**Lemma 1.** Given A1 and A2, the following holds:

$$\frac{dx_2^*}{dx_1^*} \leq 1 \iff r_1(x_1^*) \leq r_2(x_2^*)$$

*Proof.* Let  $A$  and  $B$  be defined as in P6. By (1) of P6, it is obvious that  $u_1''(x_1^*)B \geq u_2''(x_2^*)A \iff \frac{dx_2^*}{dx_1^*} \leq 1$ . Therefore, it suffices to prove

$$u_1''(x_1^*)B \geq u_2''(x_2^*)A \iff r_1(x_1^*) \leq r_2(x_2^*).$$

By the definition of  $i$ 's Arrow-Pratt coefficient,  $u_1''(x_i) = -r_i(x_i)u_i'(x_i)$ . Thus,

$$u_1''(x_1^*)B - u_2''(x_2^*)A = -r_1(x_1^*)u_1'(x_1^*)B + r_2(x_2^*)u_2'(x_2^*)A.$$

By P3,  $u_1'(x_1^*)B = u_2'(x_2^*)A$  holds and, by A1 and P4, this value is positive. Let  $u_1'(x_1^*)B = u_2'(x_2^*)A = C > 0$ . Then, we obtain

$$-r_1(x_1^*)u_1'(x_1^*)B + r_2(x_2^*)u_2'(x_2^*)A = (r_2(x_2^*) - r_1(x_1^*))C.$$

Since  $C$  is positive, the sign of  $u_1''(x_1^*)B - u_2''(x_2^*)A$  is the same as that of  $r_2(x_2^*) - r_1(x_1^*)$ .  $\square$

This lemma says that if player 1 is more risk tolerant (resp., risk averse) than player 2 at some point on the locus, the former gets more (resp., less) than the latter from the small increase  $\Delta M$  in the bargaining pie.

Building on Lemma 1, we identify the relationship between risk attitude and the (in)equality bias of the Nash allocation, where the bias is evaluated in relation to EA and NNA. Before presenting this result, let us assume for analytical simplicity that  $u = u_1 = u_2$  and  $v_1 < v_2$ ; in other words, we assume that utilities are identical and that player 1 is in a weaker bargaining position than player 2 in terms of the disagreement outcome.<sup>8</sup> Under this assumption, one easily verifies that

$$NNA_1(M, v) = \frac{M + v_1 - v_2}{2} < \frac{M}{2} = EA_1(M, v)$$

holds true.

**Theorem 1** (Equality bias theorem). Suppose  $u = u_1 = u_2$  and  $v_1 < v_2$ . The following statements hold.

- 1 (Equality bias). If  $u$  exhibits IARA, then

$$NNA_1(M, v) < x_1^*(M, v) < EA_1(M, v)$$

- 2 (No bias). If  $u$  exhibits CARA, then

$$NNA_1(M, v) = x_1^*(M, v)$$

- 3 (Inequality bias). If  $u$  exhibits DARA, then

$$x_1^*(M, v) < NNA_1(M, v) < EA_1(M, v)$$

*Proof.* We first prove the second statement. If  $u$  exhibits CARA, then for any  $(x_1^*, x_2^*)$  on the locus,  $r_1(x_1^*) = r_2(x_2^*)$  holds, which means that  $\frac{dx_2^*}{dx_1^*} = 1$  by Lemma 1. Thus, the locus in this case becomes a 45-degree line starting from the point  $(v_1, v_2)$ , and thus, it coincides with the locus of NNA.

Next, we prove the first statement. Suppose that  $u$  exhibits IARA. Take any  $(x_1^*, x_2^*)$  on the locus that is sufficiently close to  $(v_1, v_2)$ . Then,  $x_1^* < x_2^*$  holds from continuity of the locus (by P5). IARA of  $u$  implies that  $r_1(x_1^*) < r_2(x_2^*)$ , which together with Lemma 1 implies that  $\frac{dx_2^*}{dx_1^*} < 1$ . Therefore, as long as  $x_1^* < x_2^*$ , we have  $\frac{dx_2^*}{dx_1^*} < 1$ . In addition, as  $x_1^*$  approaches  $x_2^*$ ,  $\frac{dx_2^*}{dx_1^*}$  also

---

<sup>8</sup>One might argue that the assumption of utilities being identical is restrictive, because two bargainers often have different risk attitudes; for example, in the context of bargaining between a firm and a labor union, the former is typically risk-neutral, while the latter is risk-averse (McDonald and Solow, 1981). However, we note that such a case with different risk attitudes is inside our scope, because the difference could arise from different points on the same utility function.

approaches 1. It follows that  $x_1^*$  is never larger than  $x_2^*$ , which implies  $x_1^*(M, v) < EA_1(M, v)$ . Since  $x_1^* < x_2^*$  always holds on this locus, the locus of NA is located in the lower region of that of NNA, meaning that  $NNA_1(M, v) < x_1^*(M, v)$ .

The proof of the third statement is similar to that of the first statement. Take any  $(x_1^*, x_2^*)$  on the locus that is sufficiently close to  $(v_1, v_2)$ . Then,  $\frac{dx_2^*}{dx_1^*} > 1$  holds by DARA of  $u$  and Lemma 1. Thus, the difference between  $x_2^*$  and  $x_1^*$  becomes larger on the locus and thus  $x_2^* > x_1^*$  holds true for any point on the locus. Since  $\frac{dx_2^*}{dx_1^*} > 1$  holds true for any point on the locus, the locus is located in the upper region of that of NNA, meaning that  $x_1^*(M, v) < NNA_1^*(M, v)$ .  $\square$

The global property on the risk attitude has a crucial role on determining which kinds of bias exists in a bargaining problem. Especially, if utilities are assumed to exhibit DARA, the initial inequality between the bargainers becomes widens through bargaining. It is commonly believed that the utility function of a firm is DARA because it becomes more risk tolerant as its size becomes larger. If this presumption is valid, our results indicate that their bargaining amplifies the inequality between the strong and the weak.

Our results also have implications for the interpretation of bargaining outcomes in the experimental literature. Existing studies typically interpret equitable agreements among asymmetric players as a result of norms, entitlement, fairness considerations, etc. However, this theorem suggests that the bias towards the equal split is explained by Nash bargaining theory without these behavioral factors. In later sections, we will discuss this point in more detail.

To better understand the size of the bias, we apply Theorem 1 to particular forms of utility functions. We define

$$u_i(x_i) = \begin{cases} \frac{x_i^{1-b_i}}{1-b_i} & \text{if } b_i \neq 1, \\ \log(x_i) & \text{if } b_i = 1 \end{cases}$$

Note that this function depends on parameter  $b_i$ .<sup>9</sup> It is known that this utility function is DARA and risk-averse when  $b_i > 0$ ,  $u_i(x_i) = x_i$  when  $b_i = 0$ , and IARA and risk-loving when  $b_i < 0$ . The trajectory of the Nash allocations under this utility function is depicted in Figure 1. This shows that the shape of utility functions has a significant impact on the bargaining outcome. Thus, if we specify utility functions to be  $u_i(x_i) = x_i$  ignoring the true form, we might obtain a wrong conclusion.

Finally, we present some extensions of Theorem 1.  $u_1$  is said to be *more risk averse than*  $u_2$  if there exists a positive concave function  $\psi$  such that  $u_1(x) = \psi(u_2(x))$  for all  $x$ .<sup>10</sup>

**Corollary 1.** Suppose  $v_1 < v_2$ . The following statements hold true.

- 1 (Equality bias). If  $u_1$  and  $u_2$  satisfies IARA and  $u_2$  is more risk-averse than  $u_1$ ,

$$NNA_1(M, v) < x_1^*(M, v) < EA_1(M, v)$$

- 2 (Inequality bias). If  $u_1$  and  $u_2$  satisfies DARA and  $u_1$  is more risk-averse than  $u_2$ ,

$$x_1^*(M, v) < NNA_1(M, v)$$

In the literature on axiomatic bargaining theory, it is well known that being risk-averse puts the person in a worse bargaining position (Roth, 1979; Rausser and Simon, 2016). Incorporating these results into our framework will allow us to gain some insight into the negotiations that occur when one player is risk-averse and the other is risk-loving.

<sup>9</sup>This utility function is known to exhibit *constant relative risk averse*.

<sup>10</sup>A function  $\psi$  is *positive* if  $\psi(x) > 0$  for all  $x$ .

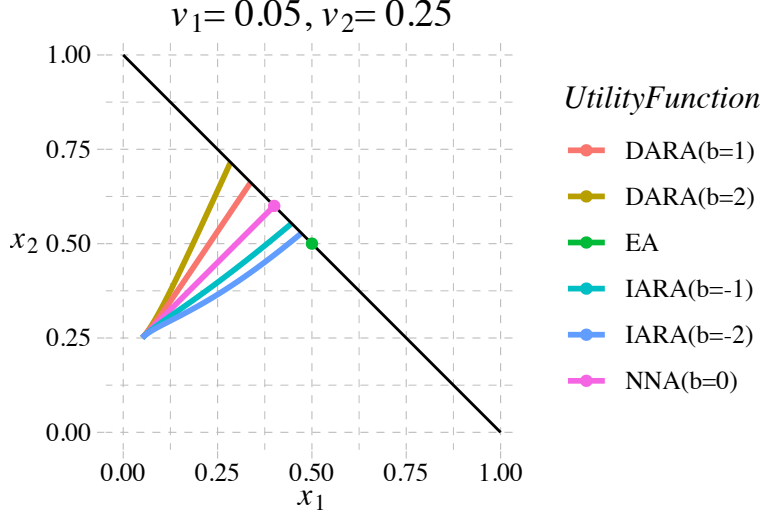


Figure 1: The size of bias

**Corollary 2.** Suppose  $v_1 = v_2 = 0$ . If  $u_1$  is risk-averse in the sense that  $u'' < 0$  for all  $x_1$  and  $u_2$  is risk-loving in the sense that  $u'' > 0$  for all  $x_2$ , then  $x_1^*(M, v) < x_2^*(M, v)$  for all  $M > 0$ .

It is well known that people become risk-loving in a loss domain. Thus, the situation in this corollary happens when one player is in the gain domain and the other in the loss domain, which is evident in various settings, e.g., in the repayment of debts from corporations to banks or in territorial disputes in international conflicts. However, the result in this corollary has been overlooked in previous studies because they have focused only on risk-averse individuals ( $u_i'' < 0$ ).

#### 4. Bargaining model with a reference dependent utility

This section develops a behavioral bargaining theory wherein entitlement, reference, and fairness concerns are taken into account. Various studies in the past have pointed out that behavioral factors other than the size of bargaining pies and the options at the breakdown affect their negotiation outcomes in experiments and real-life situations. For example, Kahneman et al. (1986) point out that the association with past deals acts as a reference point in current negotiations. It has also been argued that people's sense of entitlement to the bargaining pie (Hoffman and Spitzer, 1985; Hoffman et al., 1994), norm (Meyer, 1992; Andreoni and Bernheim, 2009), and the preference for equality and fairness (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) affect people's behavior.

The new theory developed here adopts a reference-dependent utility à la Tversky and Kahneman (1979) and Köszegi and Rabin (2006). In our terminology, we assume that people's sense of entitlement regarding bargaining pies and disagreements forms the reference point. In other words, a bargaining agreement that falls short of the entitlement is an outcome that belongs to the loss domain, while an outcome that exceeds the entitlement belongs to the gain domain. A player could have a stronger passion for satisfying the entitlement than gaining more if we model a loss aversion. Since such asymmetry is usual even when we interpret the reference as the point to which norms, status quo, fairness, and equality appeal, our model encompasses these concepts from a mathematical perspective.

It is possible to construct a unified framework for behavioral bargaining theory by capturing the impact of entitlement in a utility function. Our setup developed in the previous sections is

sufficiently flexible in dealing with the asymmetry between loss and gain, for instance, risk-loving in the loss and risk-avoidance in the gain domain. In addition, the results about the (in)equality bias are key to understanding when and how the sense of entitlements affects their bargaining outcomes. This unified theory builds on existing research findings and allows us to understand the results of various negotiation experiments, which have been challenging to discuss in a unified manner within a single framework.

Consider a bargaining between two players having entitlements  $E_1 \geq 0$  and  $E_2 \geq 0$ . We consider a reference dependent utility wherein the entitlement has a role of the reference. One remark is that the reference point is not the origin (the point of  $x_1 = x_2 = 0$ ) because we set the origin as the worst case they imagine. Therefore, the origin can be different from the disagreement point in the negotiation.

This setting allows us to describe the details of a negotiation situation more accurately than existing models. For example, consider the following two situations. In one case, an unemployed person receives an excellent joint venture offer from a friend just as he has obtained a job at a firm with a fixed wage of  $w$ , and he must negotiate with the friend for his share by using  $w$  as an outside option. In the second case, the unemployed person is replaced by an employee already working for a fixed wage  $w$ ; the other conditions are the same as in the first case. Leaving out subtleties such as future risk and uncertainty, transaction costs, and time discounting, in the traditional bargaining theory, the situation is the same for the first and second cases in the sense that there is a bargaining pie with a disagreement payoff of  $w$ . In contrast, our model treats the two cases differently. In the first case, the origin and the reference points are 0, which is strictly worse than the disagreement payoff  $w$ , and in the second case, the origin is 0 and the reference and disagreement points are  $w$ . These differences may result in different agreements during the negotiation phase.

We model the utility function as a monotonically increasing function considering asymmetry before and after the reference point. Let  $\Psi$  be such a function on  $[0, \infty)$  with  $\Psi(0) = 0$ ,  $\Psi'(x) > 0$ , and  $\Psi''(x) < 0$ . Then, the utility function is defined as the result of its simple transformation to model a loss aversion. Given a loss aversion parameter  $\lambda_i \geq 1$ , for any  $x_i \geq 0$ ,

$$u_i(x_i; E_i) = \begin{cases} \Psi(x_i - E_i) & \text{if } x_i \geq E_i, \\ -\lambda_i \Psi(E_i - x_i) & \text{if } x_i < E_i. \end{cases} \quad (2)$$

If needed, it is possible to add some positive constant in order to ensure that  $u_i(x) \geq 0$ . If we don't think about the change of  $E_i$  explicitly, we just write  $u_i(\cdot)$  instead of  $u_i(\cdot; E_i)$ .

We call  $\Psi$  a component function of utility function  $u_i$ . Figure 2 visualize the reference-dependent utility functions when its component function is IARA or DARA. A detailed analysis about IARA and DARA functions are found in Appendix 2.

## 5. Effect of the equal split norm on the bargaining outcome

In this section, we apply our behavioral bargaining model to explain the equality bias, which is quite different from the discussion in Section 3. Here, we consider the effect of a norm for the equal split that experimental participants initially have from their daily lives and use to deal with the experimental bargaining situation. Such an explanation for the equality bias is widely accepted among experimental economists, but its theoretical foundations have never been developed. The reason why the equal split is evident in our society has been discussed by evolutionary bargaining theory (Young, 1993; Binmore et al., 2003; Ellingsen, 1997) as well as our companion paper (Kamijo, 2023b) using the framework of the behavioral bargaining theory.

We assume that players have the equal split norm in a simple bargaining problem  $(M, v)$ . We

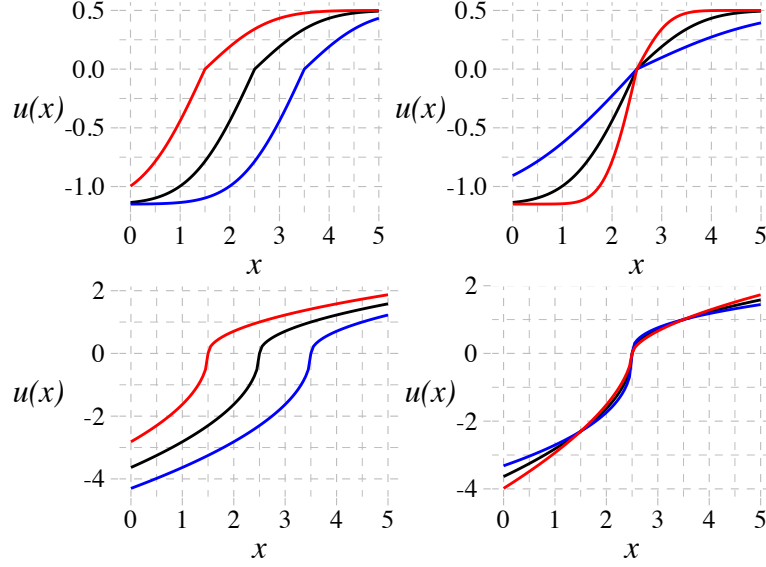


Figure 2: Concave IARA (top two panes) and DARA functions (bottom two panes). Left panes correspond to different location parameters and right panes to different shape parameters

set the entitlements to be equal to the equal split  $(M/2, M/2)$  and derive the Nash allocation. It turns out that the allocation agrees with bargaining outcomes observed in the laboratory.

To obtain a sharp result about the effect of the equal split norm on the bargaining outcome, we assume that a component function is neutral (i.e.,  $\Psi(x) = x$ ) in this section. Thus, the utility function is simplified as follows.

$$u_i(x_i; E_i) = \begin{cases} x_i - E_i & \text{if } x_i \geq E_i, \\ -\lambda_i(E_i - x_i) & \text{if } x_i < E_i. \end{cases}$$

If we consider the equal split norm,  $E_i$  in the above definition should be  $M/2$ .

The Nash allocation in this case is obtained as the solution to the following maximization problem

$$\max_{v_2 \leq x \leq M-v_1} (u_1(M-x; M/2) - u_1(v_1; M/2)) \times (u_2(x; M/2) - u_2(v_2; M/2)).$$

By solving the above maximization problems under  $v_1 \leq v_2$ , we obtain the following result.

**Theorem 2** (Equality bias caused by the equal split norm). Under  $\Psi(x) = x$ ,  $E_i = M/2$  for  $i = 1, 2$  and  $v_1 \leq v_2$ , the Nash allocation is given as follows:

- (A) If  $\lambda_2 v_2 - v_1 \leq (\lambda_2 - 1) \frac{M}{2}$  and  $v_1 \leq v_2 \leq M/2$ ,

$$x_1^{ENNA}(M, v) = x_2^{ENNA}(M, v) = M/2.$$

- (B) If  $\lambda_2 v_2 - v_1 > (\lambda_2 - 1) \frac{M}{2}$  and  $v_1 \leq v_2 \leq M/2$ ,

$$x_1^{ENNA}(M, v) = \frac{1 + \lambda_2}{4} M + \frac{v_1}{2} - \frac{\lambda_2 v_2}{2}, \quad \text{and} \quad x_2^{ENNA}(M, v) = \frac{3 - \lambda_2}{4} M - \frac{v_1}{2} + \frac{\lambda_2 v_2}{2}.$$

- (C) If  $v_1 \leq M/2 \leq v_2$  and  $v_1 + v_2 \leq M$ ,

$$x_1^{ENNA}(M, v) = \frac{1}{2}M + \frac{v_1}{2} - \frac{v_2}{2}, \quad \text{and} \quad x_2^{ENNA}(M, v) = \frac{1}{2}M - \frac{v_1}{2} + \frac{v_2}{2}.$$

*Proof.* Because of the property of the Nash solution and the assumption of linearity, the maximization problem can be transformed to

$$\max_{0 \leq x \leq M/2 - v_1} (\lambda_1 M/2 - \lambda_1 x - u_1(v_1; M/2)) \times (\lambda_2 M/2 + x - u_2(v_2; E/2))$$

where  $x \in [0, M/2 - v_1]$  is the player 2's increase in its share from  $M/2$ . In addition,  $u_i(v_i; M/2)$  varies due to the relation of  $M/2$  and  $v_i$ . If  $v_i \leq M/2$ ,  $u_i(v_i; M/2) = \lambda_i v_i$ , and if  $v_i > M/2$ ,  $u_i(v_i; M/2) = \lambda_i M/2 + (v_i - M/2)$ .

Solving the maximization problem, we obtain the result.  $\square$

We refer to the allocation in this theorem as the *Egalitarian Neutral Nash Allocation (ENNA)*.

The ENNA provides a fruitful view of the equality bias caused by the equal split norm and helps understand the bias observed in the laboratory (Anbarci and Feltovich, 2013, 2018; Birkeland and Tungodden, 2014). The ENNA becomes the equal split allocation (the egalitarian solution) when  $v_1$  and  $v_2$  are relatively small and not so different (Regions  $A$  and  $A'$  of the left panel of Figure 3), it becomes the neutral Nash allocation (the standard solution) when one of  $v_1$  and  $v_2$  is more than  $M/2$  (Regions  $C$  and  $C'$ ), and it becomes the intermediate values between the equal split and the neutral Nash allocation when  $v_1$  and  $v_2$  are moderate in size and not so similar (Regions  $B$  and  $B'$ ). Therefore, the ENNA indicates that the equality bias happens only when the disagreement outcome of the stronger player is less than half of the pie, and its effect is more evident when it is smaller.

The left panel of Figure 3 is also helpful to understand how the loss aversion parameters affect the bargaining outcome. When the  $\lambda$ 's of the two players are large, almost any bargaining situation wherein their disagreement payoff is less than the half of a pie goes to the equal split (since the area  $A + A'$  expands and  $B + B'$  shrinks). In contrast, when the  $\lambda$ 's are small, only the almost symmetric situations ( $v_1$  and  $v_2$  are close each other) leads to the equal split.

To understand the discontinuous effect of the disagreement outcome, let's see the right panel of Figure 3. This shows how the share of player 2 varies according to  $v_2$  with fixing  $v_1 = 0$  (the black solid line). The red dashed line corresponds to the standard solution and the blue dashed one corresponds to the constrained egalitarian solution.<sup>11</sup> Thus, it is apparent that the ENNA is always between these two solutions. In fact, the ENNA coincides with the equal split when  $v_2$  is small and it coincides with the standard solution when  $v_2 \geq M/2$ . Notice that the region (B) connecting the egalitarian and the standard solutions shrinks as  $\lambda_2$  becomes larger. In that case, around  $M/2$ , ENNA shifts sharply from the egalitarian solution to the standard solution, which would appear to be almost discontinuous at this point (not in the mathematical sense, but in terms of verification from the data).

From this discontinuity of the agreed outcome predicted by ENNA, we can hypothesize that the disagreement outcome has a weak influence when it is less than half of the pie and has a strong influence when it is more than half. This hypothesis was actually tested in the experiment of Anbarci and Feltovich (2018). In their study, experimental participants bargain over the distribution of the pie in two bargaining formats: the Nash demand game and the unstructured bargaining, where bargainers can freely bargain for the division of the pie during a given time

<sup>11</sup>A constrained egalitarian solution refers to a solution that aims for the equality as much as possible within the constraints imposed by the disagreement outcomes.

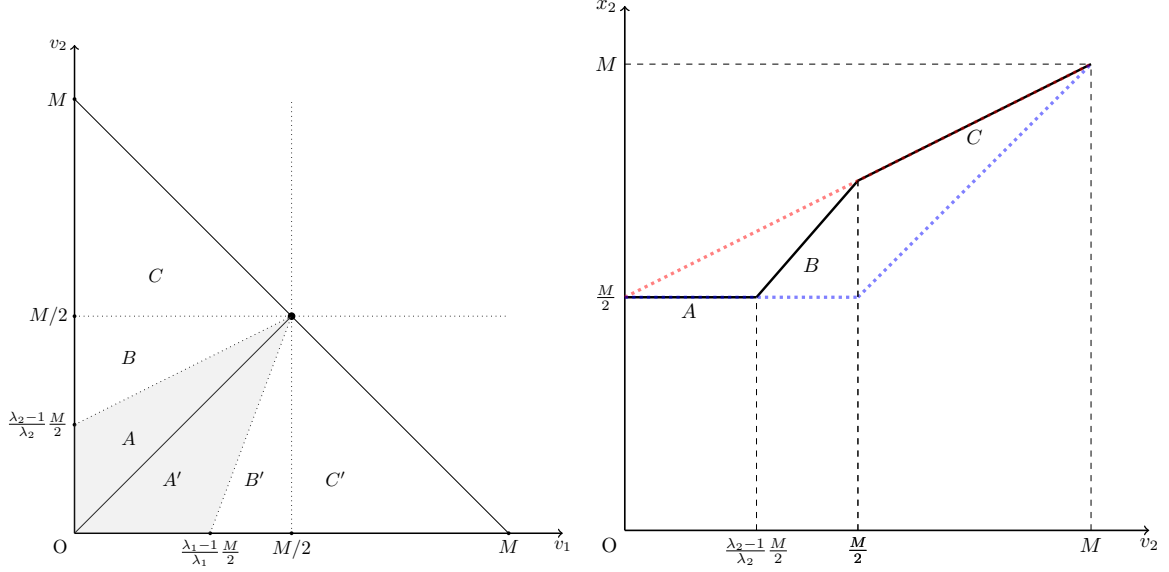


Figure 3: Three regions described by ENNA. Regions A, C and B correspond to egalitarian solution, the equal surplus solution, and their convex combination, respectively (left panel). Right panel describes how the share of player 2 varies as  $v_2$  increases with fixing  $v_1 = 0$ .

interval. They define the measure of how the bargaining agreement is sensitive to the change in the disagreement outcome as

$$S(M, v) = \left| \frac{\partial x_1}{\partial v_1} \right| + \left| \frac{\partial x_1}{\partial v_2} \right|.$$

They calculated these values from their data in the unstructured bargaining experiment and reported that this is 0.675 when  $v_i \leq M/2$  for  $i = 1, 2$ , and it is 0.870 when one of  $v_i$  is greater than  $M/2$ . Such discrepancy is much evident in the experiment of the Nash demand game (the former is 0.254 and the latter is 0.723).<sup>12</sup> Anbarci and Feltovich (2018) argued that this discontinuity is the result of switching the focal point; when the equal split is the individually rational, it becomes the focal point, but when it is not, the standard solution seems to become focal. ENNA, defined by the Nash solution with an equal split norm, captures this phenomenon (see the right panel of Figure 3), and the discontinuity naturally arises without considering the switch of the focal points. In other words, ENNA gives a theoretical justification for why such switch of the focal points occurs.

Another merit of the ENNA is that it captures the distinction between the bargaining outcome influenced by the equal split norm and one caused by the preference for equality. If the bargaining outcome results from the preference for equality, it should be like the constrained egalitarian solution (the blue line in the right panel of Figure 3). The ENNA and the constrained egalitarian solution coincide when  $v_2$  is small but become different as  $v_2$  becomes large. Especially, when  $v_2 > M/2$ , these two solutions provide quite different allocations.

## 6. Applications

### 6.1. Bargaining environment

The discussion in the previous section shows that our reference-dependent utility bargaining model nicely describes bargaining outcomes under the influence of the equal split norm. In this section, we apply the BBT to other behavioral patterns that are often observed in the

<sup>12</sup>Kamijo (2023c) shows that the risk dominant equilibrium of Nash demand game coincides with the Nash bargaining solution for very wide class of utility functions.

experimental literature.

We call  $(M, v, (E_1, E_2))$  a *bargaining environment*. Since our reference-dependent utility function  $u_i$  depends on entitlement  $E_i$ , the Nash allocation calculated from the reference dependent utility functions depends not only the bargaining problem  $(M, v)$  but also the entitlement vectors  $(E_1, E_2)$ . As a result, it becomes a function that associates any bargaining environment  $(M, v, E_1, E_2)$  with the efficient allocation. We denote the Nash allocation by

$$x^*(M, v, (E_1, E_2))$$

to emphasize the dependency on  $E_1$  and  $E_2$ .

In the following subsections, we hold the view that, whether it is intentional or not, different experimental manipulations generate different bargaining environments, even though the bargaining problem is the same.

Before going to the application parts, we reexamine the relationship between utility function  $u$  defined in (2) and the component function  $\Psi$  in terms of risk attitude. One easily verifies that IARA/DARA of the component function  $\Psi$  is inherited to  $u$  in the loss and gain domains. Let  $r_\Psi$  be the Arrow-Pratt coefficient of a function  $\Psi$ . Then, after some calculation, in the gain domain ( $x_i > E_i$ ), we have  $r_i(x_i) = r_\Psi(x_i - E_i)$ . In contrast, in the loss domain ( $x_i < E_i$ ), since  $(\Psi(E_i - x_i))' = -\Psi'(E_i - x_i)$  and  $(\Psi(E_i - x_i))'' = \Psi''(E_i - x_i)$ , we have  $r_i(x_i) = -r_\Psi(E_i - x_i)$ . Therefore,  $r'_i(x_i) = r'_\Psi(x_i - E_i)$  in the gain domain and  $r'_i(x_i) = -(-1)r'_\Psi(E_i - x_i) = r'_\Psi(E_i - x_i)$  in the loss domain. Thus, in both cases, IARA (DARA) of  $\Psi$  is succeeded to  $u_i$  defined in (2). A detailed analysis about IARA and DARA utility functions are provided in Appendix 2.

We use the terminology of IARA (DARA) model to refer to a situation that both bargainers have the same concave IARA (DARA) function  $\Psi$  and their utility is defined by (2).

## 6.2. Bargaining for “manna from heaven”

As the first application of our bargaining model, let us consider a typical practice of traditional bargaining experiments. Participants in the bargaining experiments are often instructed as follows:

You and the person matched to you bargain over a £20.00 prize. You do this by sending and receiving proposals for dividing the prize during a “negotiation stage” of the game. Below is an example of how the bottom portion of your computer screen will look during the negotiation stage.

...

[ **The decision screen are shown and the explanation about how they can make a proposal, accept one of the proposals, and end the negotiation stage is continued** ]

...

If you or the other person ends the negotiation stage early, or if the time available for proposals ends without you reaching an agreement, then you receive an “outside option”, and the other person receives a different “outside option”. These outside options are chosen randomly by the computer, and vary from round to round and from person to person. In each round, you and the person matched to you are informed of both of your outside options at the beginning of the negotiation stage.

[from an experimental instruction of Anbarci and Feltovich (2013). The bold-font text is inserted by the authors of the current study]

In this situation, both the disagreement payoffs and the bargaining pie are windfall gains in the sense that they are not related to the participants’ skill, status, efforts and actions. Then, it is reasonable to think that players do not feel any entitlement to a bargaining surplus or disagreement payoffs. One possible model of capturing this situation is to assume that they are under the influence of the equal split norm and thus think  $E_1 = E_2 = M/2$ . This situation is already analyzed in the previous section and we observed that the equality bias naturally happens. Another way to model this situation is to naively assume that  $E_1 = E_2 = 0$  holds. In this subsection, we analyze the latter case.

Suppose that  $E_1 = E_2 = 0$  holds. Thus, the bargaining environment in this case is  $(M, v, (0, 0))$ .

If the disagreement point  $v$  is different from the origin but they do not feel any entitlement to their disagreement point either, the following result holds by our Theorem 1.

**Proposition 1.** Suppose that the bargaining environment is  $(M, v, (0, 0))$ . Then, the equality bias occurs in the IARA model, no bias occurs in the CARA model, and the inequality bias occurs in the DARA model.

Using Proposition 1, we offer new interpretations of existing results in the laboratory. The IARA model is consistent with lots of experimental evidence that point out the existence of the equality bias.

Some studies examine the equality bias from a different angle. Anbarci and Feltovich (2013) experimentally observe that subjects’ agreement is less responsive to changes in disagreement outcomes compared to the neutral Nash solution’s prediction. Because the equal split is utterly unresponsive to disagreement points, this phenomenon is entirely the flip side of the coin of the equality bias. In fact, as a similar result to our Theorem 1, Kamijo (2023a) shows that the outcome of the Nash solution is less responsive to disagreement outcomes when the utility function satisfies IARA.

Experiments about dictator and ultimatum games have revealed that experimental participants show a preference for fairness and equality (Knez and Camerer, 1995; Hoffman and Spitzer, 1985; Hennig-Schmidt et al., 2018). So, the equality bias observed in a unstructured bargaining experiment can be explained from this preference.<sup>13</sup> Our Proposition 1 provides another explanation for this phenomenon. Without specific fairness consideration in a utility function, the IARA utility function can explain the equality bias when the bargaining pie and disagreement points are given to players as “manna from heaven.”

### 6.3. Bargaining with earned disagreement outcomes

The existing literature has documented that the equality-biased outcomes often observed in bargaining experiments arise from the lack of entitlements that participants feel to the bargaining pie and the outside option. To manipulate the feelings of entitlements, one of the most common approaches is to introduce a production stage wherein participants can earn money or experimental tokens by conducting some tasks (Baranski, 2016, 2019; Cappelen et al., 2007, 2011). A common observation from the bargaining experiment with production stage is that the 50–50 split frequently observed in the ultimatum and dictator games without production is less frequent in experiments in which the participants produce the surplus on their own. In an unstructured bargaining experiment, Luhan et al. (2019) found that if a bargaining pie is the product of their effort and performance, the agreement other than the equal split is often justified. Takeuchi et al. (2022) also found that if there is a production stage, the egalitarian

---

<sup>13</sup>Birkeland and Tungodden (2014) explained the equality bias in the unstructured bargaining experiment by incorporating concerns for fairness into the utility functions and applying the Nash solution.

allocation is less observed and allocations close to other familiar solutions are observed more frequently.

In an unstructured bargaining experiment with a production stage, the earned outcome in the production stage is then used as the input in the subsequent bargaining stage like the following instruction:

You and your partner can use the earnings from Part 1 [**the production stage**] and engage in a joint investment that gives to you two a joint profit, which is higher than the sum of earnings from Part 1. The task for you in Part 2 will be to negotiate with your partner in order to reach an agreement on how to divide the joint profit.

The joint profit depends on the amount of money invested. The relationship between the money earned in Part 1 and the joint profit is as follows: Joint profit = Your earnings + your partner's earnings +  $\alpha$ , where  $\alpha$  is a number that changes from round to round. Note that when you make a joint investment, you have to invest all the money earned in Part 1.

[from an experimental instruction of Takeuchi et al. (2022). The bold-font text is inserted by the current authors]

As in the above instruction, suppose that the disagreement outcome is the one that experimental participants earn through some real effort tasks. They can obtain an extra surplus by putting their disagreement payoffs into the opportunity of a joint project. In this case, it is reasonable to assume  $E_i = v_i$  for  $i = 1, 2$ . Thus, the bargaining environment becomes  $(M, v, (v_1, v_2)) = (M, v, v)$ .

In our setting wherein the heterogeneity between players are captured by the different reference points, the utility generated by the increment from the reference point is precisely the same for both. Hence, from Lemma 1, the slope of the bargaining trajectory continues to be 1. Thus, we have the following result.

**Proposition 2.** Suppose that the bargaining environment is  $(M, v, v)$ . In any of IARA, CARA or DARA models, the bargaining trajectory coincides with that of NNA.

This proposition says that the neutral Nash solution appears when  $E_i = v_i$  for  $i = 1, 2$ . This prediction is supported from the experiment of Takeuchi et al. (2022). They manipulate how the bargaining surplus is related to their input (disagreement payoff) earned in the production stage. No relationship is mentioned in the baseline, and it is explained in the constant surplus treatment that the bargaining pie is the sum of the fixed surplus and the subjects' input (i.e.,  $M = \alpha + v_1 + v_2$  with  $\alpha > 0$ ). They classified the agreed outcomes into nine categories based on the three well-known solutions, the equal-split, the neutral Nash solution, and the proportional solution, and they found that the most frequent agreement in these two treatments is the category close to the neutral Nash solution.

Another example of the disagreement point being the entitlement is the experiments on price bargaining between a seller and a buyer wherein a price below the seller's reservation price (above the buyer's reservation price) entail a negative payoff to that person. In this situation,  $E_i = v_i = 0$  for  $i = 1, 2$  holds and the neutral Nash solution is the equal split. Raiffa (1982) observed that "the obvious focal point would be in the middle (...), and that's what happens overwhelmingly in experimental negotiations" (Raiffa 1982, p52). This result is also consistent with Proposition 2.

#### 6.4. Entitlement paradox in loss domain bargaining

Incorporating entitlements into the model enables us to consider the problem of how to share the surplus in the loss domain or share the cost among the stakeholders. In the literature

on unstructured bargaining experiments, bargaining in the loss domain is done by forcing participants into a loss frame, as in the following experimental instruction.

In the experiment you and your partner act in the role of a head of department in a firm. Imagine that in this firm there is a total budget of 2490 points for your and your colleague's salary. In the past the policy of the firm was to pay the salary according to performance. (How the performance is determined in the experiment will be explained below.) The head of department with the higher performance was paid a salary of 1660 points and the head of department with the lower performance was paid a salary of 830 points. There is now the possibility that, due to bad economic conditions for the firm the top management is forced to cut back the salary budget, with the consequence that the hitherto valid salary claims cannot be satisfied anymore. The new total salary budget then amounts to 2050 points.

[from an experimental instruction for participants of Gächter and Riedl (2005)]

In the above explanation, if the economic situation is bad, the status quo salary will not be covered, in which case the participant will have to negotiate regarding the distribution of an amount less than the total of the status quo salary. A similar situation arises when the remaining assets of a bankrupt company are less than the total amount claimed by its creditors (Aumann and Maschler, 1985; O'Neill, 1982). The bankruptcy problem and other cost-sharing problems have been studied intensively through the axiomatic approach combined with cooperative game theory (for a survey, see Thomson (2003)).

Consider a bankruptcy problem or a claim problem wherein each player  $i$  has a claim  $c_i$  to the bankrupt firm and the remaining asset of the firm is  $M$  with  $M < c_1 + c_2$ . They have to decide how to share the remaining assets  $M$ . Several kinds of allocation rules have been proposed and investigated from a normative point of view. Almost all rules draw a trajectory through the *bargaining parallelogram*, where the outcome lies between the constrained equal award and constrained equal loss rules (see Figure 3 Panel (A)). In the panel, the blue bottom, the red top, the black and the central dashed lines are the constrained equal award, the constrained equal loss, the proportional and the Talmudic rules,<sup>14</sup> respectively).

Here, instead of applying allocation rules directly, we assume that the cost allocation is determined through the bargaining between the stakeholders. Thus, we apply our behavioral bargaining theory to this class of problems and try to know whether they agree on the allocations that a normative rule should suggest. A variety of trajectories will be obtained as the result of the variation of utility functions, not by allocation rules.

Since any shortage of the claim can be seen as the loss for that player, it is natural to set  $c_i = E_i$  for  $i = 1, 2$ . In addition, we assume  $x_i^*(M, v) \leq E_i$  because of the nature of the problem.<sup>15</sup>

We obtain the following proposition, indicating that IARA is more consistent with the normative solution concepts and DARA leads to a paradoxical result.

**Proposition 3.** Suppose that the bargaining environment is  $(M, (0, 0), (c_1, c_2))$ , where  $c_1 + c_2 > M$  and  $c_1 < c_2$ . The following statements hold true:

- (i) In the IARA model, the locus of bargaining in the bankruptcy problem always belongs to the bargaining parallelogram,

<sup>14</sup>In two-player case, the Talmudic rule coincides with the Shapley value, the nucleolus, and other normative solutions

<sup>15</sup>Without this assumption, we have to slightly modify the statement in the next proposition since when there is a large difference between  $E_1$  and  $E_2$ , it might happen that a person with smaller entitlement would obtain the share more than the entitlement as the result of the bargaining.

- (ii) In the CARA model, the locus of bargaining in the bankruptcy problem coincides with the one of the constrained equal award rule,
- (iii) In the DARA model, the locus of bargaining in the bankruptcy problem deviates from the bargaining parallelogram.

*Proof.* (i) and (ii). First, it should be mentioned that different values of  $\lambda_i$  for  $i = 1, 2$  do not change the bargaining outcomes. This is because since the share of player  $i$  is always in the loss domain,  $\lambda_i$  becomes a multiplier of the Nash product, and thus, the maximizer of the Nash product is invariant from the value of  $\lambda_i$ .

Second, by Lemma 2 in the Appendix, in the IARA model, assumption A2 is automatically satisfied, so the locus is guaranteed to be an upward-right sequence. Also, considering the locus at the origin, player 1 is more risk-averse than player 2 at the origin due to the definition of the utility function and the nature of IARA.<sup>16</sup> Therefore, from Lemma 1, the slope of the locus is greater than 1.

Furthermore, suppose that a point on the locus is in the neighborhood (lower side) of the equal loss. In this case, the degree of risk aversion of both players is almost equal, so the slope of the locus becomes closer to 1.

From the above three facts, we can see that the locus of negotiation does not deviate outside the bargaining parallelogram for IARA model.

In the case of CARA model, from Lemma 1, the slope of the locus is equal to 1 until it reaches the boundary of the bargaining parallelogram. Thus, it becomes the locus of the constrained equal gain rule.

(iii). In the DARA model, since the utility function satisfies DARA, player 1 is less risk averse than player 2 at the origin. Therefore, the slope of the trajectory becomes smaller than 1. Therefore, near the origin, the trajectory deviates from the parallelogram of negotiation.  $\square$

The second statement of the proposition shows that when the utility function exhibits CARA, it is consistent with the neutral Nash allocation. This is a generalization of Dagan and Volij (1993) showing that the Nash solution is consistent with the constrained equal gain rule when risk-neutral agents are assumed. The axiomatic analysis of bargaining solutions has revealed which ones are consistent with the allocation rule for the bankruptcy problem under the assumption of risk-neutral agents. In contrast, the proposition shows that various trajectories can be obtained by considering changes in the utility function with fixing the bargaining solutions to the Nash solution.

In the IARA model, it is possible to draw trajectories similar to those of various normative solutions, depending on the shape of the utility functions. In Panel (B1) of Figure 4, we use

---

<sup>16</sup>This is verified as follows. By the definition of a utility function, we have

$$u_i(x_i) = -\lambda_i \Psi(c_i - x_i).$$

Its Arrow-Pratt coefficient is

$$r_i(x_i) = -\frac{\lambda_i \Psi''(c_i - x_i)}{-\lambda_i \Psi'(c_i - x_i)} = -\left(-\frac{\Psi''(c_i - x_i)}{\Psi'(c_i - x_i)}\right).$$

When  $\Psi$  is IARA,  $r_\Psi(E_i) = -\frac{\Psi''(E_i)}{\Psi'(E_i)}$  is an increasing function in  $E_i$ . Therefore, we have

$$r_1(0) = -\left(-\frac{\Psi''(c_1)}{\Psi'(c_1)}\right) > -\left(-\frac{\Psi''(c_2)}{\Psi'(c_2)}\right) = r_2(0)$$

when  $c_1 < c_2$ .

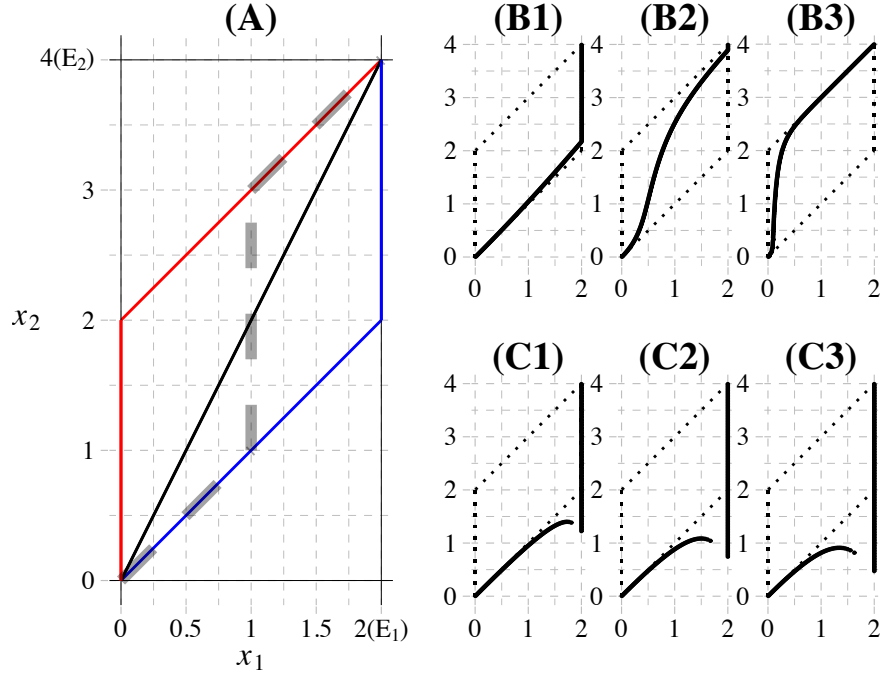


Figure 4: Pane (A) is the bargaining parallelogram. Panels (B1), (B2), and (B3) are the trajectories under IARA when the convexity is weak, medium, and strong, respectively. Panels (C1), (C2), and (C3) are those under DARA. We use pdf of a normal distribution as an IARA function and the degree of convexity is controlled by the different values of standard deviation (small/large sd correspond to strong/weak convexity in the domain). As a DARA function, we use a power-type function a la Tversky and Kahneman (1992).

the utility function that is much flatter and whose degree of convexity is weak. In such a case, we can see that the trajectory is close to the equal award rule. In contrast, when the degree of convexity is strong, the trajectory is similar to the one of the constrained equal loss rule (Panel (B3) of Figure 4). It is worth noting that, when the degree of convexity is of medium size, it is a mixture of the trajectory of the Talmudic rule and the proportional rule (Panel (B2) of Figure 4).

On the other hand, the DARA model yields inconsistent results with predictions from normative solutions. In the DARA model, not only does the trajectory deviate from the parallelogram of negotiation, but also it is discontinuous (because assumption A2 is not satisfied around the reference point). In the DARA model, as shown in Panels (C1), (C2) and (C3) of Figure 4, after a gentle rightward rise in the lower right region of the parallelogram, the locus begins to be down to the right (i.e., the share of 2 begins to decrease), and then, at some point, it jumps discontinuously to the right wall. After it reaches the right wall, the trajectory rises straight up.

As Figure 4 shows, in the DARA model, some paradoxical results happen when  $M < 2E_1$ . In this case, the bargaining share of a player with a smaller entitlement ( $E_1$ ) becomes larger than that of a player with a higher entitlement ( $E_2$ ). Empirically, no paper examines whether the entitlement paradox occurs or not. Thus, it is not easy to compare the two models from this point of view. Nevertheless, it is unlikely that the entitlement paradox is justified since at least no normative solution supports such an allocation.

Experimental investigations on the loss domain bargaining are not as abundant as those focusing on the gain domain bargaining. One exception is an experimental paper by Gächter and Riedl (2005). The experimental results of bargaining in the loss domain help us identify which of the two models is more consistent with the data. The most frequent share of the higher claim player is 65-67%, which is a slight shift of the equal loss to the egalitarian side (70.2% is the share of the higher claim player under the equal loss rule). There is no agreement on an outcome that is more favorable to the higher claim player than the equal loss rule. The second and third mode in the data is  $(E_1, M - E_1)$ , which is the allocation suggested by the constrained equal gain rule. Thus, their experimental result is in good agreement with the IARA model, which predicts that they agree on various points in the parallelogram.<sup>17</sup> In contrast, the DARA model does not have such a variation in prediction.

## 7. Data fitting

In the previous sections, we demonstrate that BBT explains important stylized facts observed in bargaining experiments by considering various types of entitlements. In this section, we empirically check the validity of BBT by using the data of Takeuchi et al. (2022).

In their experiment, an experimental participant earns  $v_i$  by a real effort task in the first stage (production stage), and then, they put  $v_i$  to a joint project and bargain for the division of the pie  $M$  in the second stage (bargaining stage). The explanation about the relationship between  $M$  and  $(v_1, v_2)$  is manipulated in their experiment: In the baseline treatment, an explicit relationship is not explained; in the constant surplus treatment, it is explained as  $M = \alpha + v_1 + v_2$ , where  $\alpha$  is the constant surplus; in the proportional surplus treatment,  $M = \beta(v_1 + v_2)$ , where  $\beta$  is a multiplication factor. Importantly, in these three treatments,  $M$  and  $(v_1, v_2)$  are the same, but only the explanations about the relationship are different. However, their manipulations are expected to affect the feelings of entitlements that experimental participants perceive, and thus, bargaining environments  $(M, v, (E_1, E_2))$  are different in the three treatments.

<sup>17</sup>In 25% percent of the agreements, a lower claim player gets more than its claim since some participants were not affected by the experimental manipulation about status quo. Their experimental setting is different from the bankruptcy problem in this respect.

The estimated model is constructed as follows. For any bargaining problem  $(M, v)$ , we can calculate the bargaining outcome if we know the type of utility function (e.g., IARA or DARA) and entitlement vectors  $E = (E_1, E_2)$ . So far, we propose several types of entitlements, each of which corresponds to a specific experimental manipulation. An entitlement function  $\gamma$  specifies the relation between a bargaining problem and the bargainers' entitlements, and  $E_i = \gamma_i(M, v)$  for  $i = 1, 2$ . To know the degree of fitness to the data, we consider the following types of entitlement functions. For each  $i = 1, 2$ ,

- Equal Split Norm:  $\gamma_i(M, v) = M/2$ .
- Manna from Heaven:  $\gamma_i(M, v) = 0$ .
- Disagreement:  $\gamma_i(M, v) = v_i$ .
- Proportion:  $\gamma_i(M, v) = \frac{v_i}{v_1 + v_2} M$

The first three are already explained in the previous section. The “Proportion” case is such that bargainers have entitlements that are proportional to their disagreements and satisfy the consistency condition ( $M = E_1 + E_2$ ).

Let  $x_{s,p}^*(M, v, E_1, E_2; \theta)$  be the agreed share for the strong bargainer (i.e., player 2 if  $v_2 > v_1$ ) of the bargaining pair  $p$ , wherein  $\theta$  is the parameter of the utility (component) function (e.g.,  $\theta = b$  when  $\Psi(x) = x^b$ ). Then, we assume that the observed  $x_{s,p}$  is

$$x_{s,p} = x_{s,p}^*(M, v, E_1, E_2; \theta) + \epsilon_p$$

where  $\epsilon_p$  is an error term and is distributed according to a normal distribution with mean  $\mu = 0$  and standard deviation  $\sigma$ .

Then, the likelihood of a data point  $x_{s,p}$  is

$$f(x_{s,p}; \theta, \sigma) = \phi\left(\frac{x_{s,p} - x_{s,p}^*(M, v, E_1, E_2; \theta)}{\sigma}\right)$$

where  $\phi$  denotes the density of a standard normal distribution.

As we explained so far,  $E_1, E_2$  are functions depending on  $M$  and  $v$ , and there are several types. Thus, we denote  $f$  by  $f^{ESN}$ ,  $f^{Manna}$ ,  $f^D$ , and  $f^{Prop}$  when the entitlement function  $\gamma_i$  is one of “Equal Split Norm,” “Manna from Heaven,” “Disagreement” and “Proportion,” and  $E_i = \gamma_i(M, v)$  for each  $i = 1, 2$ .

Given a particular type of entitlement, we calculate the log-likelihood for all data as follows:

$$LL^K(\theta, \sigma) = \sum_p \log(f^K(x_{1,p}; \theta, \sigma)),$$

where  $K$  is one of ESN, Manna, D, and Prop.

We estimate the model parameters using the data of Takeuchi et al. (2022).<sup>18</sup> To focus on the effect of the entitlements, we only consider the neutral component function  $\Psi(x) = x$ . Thus,  $\theta$ , the parameter of the component function, is fixed and not estimated. In addition, we assume that the loss aversion parameter  $\lambda$  is 2.33 for simplicity. Thus,  $\sigma$  is the only estimated variable. An estimation model using the neutral component function with a particular entitlement function is referred to as a simple behavioral model.

<sup>18</sup>In their experiments, bargaining pairs are matched randomly so that some pairs have equal disagreement payoffs. Because symmetric bargaining problems lead to an equal split of a pie theoretically and experimentally, we used data from asymmetric bargaining problems (i.e.,  $v_1 \neq v_2$ ). We performed an estimation on data where participants agreed on an efficient allocation. However, the results are much the same when inefficient deals and failed negotiations are included in the estimation.

	ESN	D (Manna)	Prop
Baseline treatment	-550.4673	<b>-459.6879</b>	-541.2521
Constant surplus treatment	-521.6796	<b>-407.8255</b>	498.4210
Proportional surplus treatment	-624.8405	-533.0858	<b>-474.1858</b>

Table 1: Log likelihood of a simple behavioral model to the three treatments of Takeuchi et al. (2022)

Note: The numbers of data are 238, 229, 229, respectively. ESN, Manna, D, Prop means "Equal Split Norm", "Manna from Heaven", "Disagreement", and "Proportion", respectively. Bold font indicates the best fit model.

The estimation results appear in Table 1. Since "Disagreement" and "Manna" perfectly coincide in the neutral case (see Proposition 2) and they both induce the outcome of the neutral Nash allocation (the equal surplus division), we only report the estimation results of "Disagreement." It shows that choosing the proper types of entitlement greatly increases fitness of the model, indicating its importance for explaining the data. This tendency is common across all treatments, but the proper model is different. "Manna from Heaven" and "Disagreement" show a good fit for the baseline and the constant surplus treatments, which indicates that the data of these treatments are close to the neutral Nash allocation. In contrast, "Proportion" best fits the proportional surplus treatment.

In addition to fitness of the data, the selected models exhibit the interpretation of the experimental manipulations. In their post-experiment questionnaire, experimental participants answered which way of division is most preferred, the equal split division, the equal surplus division, or the proportional surplus division. The results are in accordance with the selection from the simple behavioral model. They answer that the equal surplus division is most favorable in the baseline and constant surplus treatments, but the proportional surplus division is most preferred in the proportional surplus treatment.

Furthermore, we also estimate the full models of the BBT. In a full model, the parameter of the component function is also estimated. Let  $\Phi(\cdot; \mu, s)$  denote the cumulative density function for a normal distribution with mean  $\mu$  and standard deviation  $s$ . In the IARA case, we assume that  $\Psi(x) = \Phi(x; 0, s) - 0.5$  (this becomes a concave IARA function; see Appendix 2 for details). In the DARA case, we assume that  $\Psi(x) = x^b$  with  $0 \leq b \leq 1$ . In addition to  $\sigma$ , parameters  $s$  and  $b$  are estimated in the IARA and DARA cases, respectively.

The results appear in Table 2. In the baseline and constant surplus treatments, "Manna from Heaven" of DARA is the best fit, which reflects the tendency that the mode of these treatments is the neutral Nash allocation, but the distribution of agreed outcomes is skewed towards the proportional side, which is the direction of the inequality bias (see Theorem 1 and Proposition 1). In contrast, the "Proportion" of IARA shows the best fit for the proportional surplus treatment. Therefore, even though we consider the parameter of the component functions, selected entitlements are kept the same. Also, considering both elements (components and entitlement functions) generates the best fit model, but the size of the fruit is modest.

## 7. Concluding remarks

This paper presents two explanations for the equality bias observed in bargaining experiments. One is a canonical approach that assumes a standard utility function that accounts for risk attitudes. This approach justifies the equality bias if players have IARA utility functions. The other approach is the behavioral approach, which assumes that players have a certain sense of entitlement to the bargaining pie. This approach allows us to theoretically model the equal split norm, and we show that the resulting bargaining outcome, named ENNA, provides essential insights into the equality bias that the equal split norm, rather than egalitarian preferences, can

Baseline treatment				
	ESN	Manna	D	Prop
IARA	-550.4669	459.6879	-459.6879	-541.2521
DARA	-550.4670	<b>-457.3894</b>	-459.6879	-541.2476

Constant surplus treatment				
	ESN	Manna	D	Prop
IARA	-521.6792	-407.8255	-407.8255	-498.4144
DARA	-521.6793	<b>-403.3803</b>	-407.8255	-498.4144

Proportional surplus treatment				
	ESP	Manna	D	Prop
IARA	-624.8401	-533.0858	-533.0858	<b>-463.3777</b>
DARA	-624.8401	-474.6077	-533.0858	-464.7159

Table 2: Log likelihood of full BBT models to the three treatments of Takeuchi et al. (2022)  
For the explanations, please see caption of Table 1.

produce.

In addition to explaining the equality bias, the behavioral model provides many applications. BBT expands the scope of analysis from a bargaining problem  $(M, v)$  in the traditional bargaining theory to a bargaining environment  $(M, v, (E_1, E_2))$ . By considering the relation between the experimental manipulation and the type of entitlement vector  $(E_1, E_2)$ , BBT provides unified perspectives for understanding the existing experimental literature, such as bargaining for manna from heaven, bargaining based on earned disagreement outcomes, and loss domain bargaining. Furthermore, data fitting demonstrated the effectiveness of BBT, showing that selecting an appropriate type of entitlements improves the model's fitness to the data.

Extensions and future directions of BBT are as follows. For theoretical studies, by using the framework of BBT, it is expected that further theoretical analysis on the feeling of entitlements that people form proceeds. For example, a companion paper to this study examines how people form feelings of entitlement and discusses the emergence of distributive and equal split norms (Kamijo, 2023b). For experimental studies, while BBT provides predictions to bargaining environments, it remains unclear what kind of manipulation creates what kind of sense of entitlement. We hope that future experimental studies will move in the direction of clarifying the link between entitlements and experimental manipulations. For empirical studies, BBT is expected to create an elaborated data-fitting model, especially for the formation of entitlements based on experience. In the near future, the path dependence and status quo effects might be successfully modeled and explained from data using BBT.

## Appendix 1: Proof of P4, P5 and P6

### Proof of P4

*Proof.* Since the Nash allocation is obtained as an interior solution, we have  $A > 0$  and  $B > 0$ . Then, the identity equation of P3 can be rewritten as follows:

$$u'_1(x_1) = \frac{A}{B} u'_2(x_2)$$

Thus,

$$u'_1(x_1)u'_2(x_2) - u''_2(x_2)A = \frac{A}{B}(u'_2(x_2))^2 - u''_2(x_2)A = A\left(\frac{(u'_2(x_2))^2}{B} - u''_2(x_2)\right).$$

By A1 and A2 of utility functions, we have  $-u''_2(x_2) > -\frac{(u'_2(x_2))^2}{u_2(x_2)}$ . Therefore,

$$A\left(\frac{(u'_2(x_2))^2}{B} - u''_2(x_2)\right) > A\left(\frac{(u'_2(x_2))^2}{u_2(x_2) - u_2(v_2)} - \frac{(u'_2(x_2))^2}{u_2(x_2)}\right) \geq 0$$

Thus, we have  $u'_1(x_1)u'_2(x_2) - u''_2(x_2)A > 0$ . The latter part of the lemma can be shown in a similar way.  $\square$

### Proof of P5

*Proof.* We prove the uniqueness part, which together with Jansen and Tijs (1983) implies the continuity part.

If  $M = v_1 + v_2$ , then uniqueness immediately follows from P1. In what follows we assume  $M > v_1 + v_2$ . We define  $f : (v_1, M) \rightarrow \mathbb{R}$  by

$$f(x_1) = (u_1(x_1) - u_1(v_1))(u_2(M - x_1) - u_2(v_2)) \text{ for all } x_1 \in (v_1, M).$$

We prove two claims.

**Claim 1.** For any  $x_1 \in (v_1, M)$ ,  $f'(x_1) = 0$  implies  $f''(x_1) < 0$ .

*Proof.* Fix  $x_1 \in (v_1, M)$ . It holds that

$$f'(x_1) = u'_1(x_1)B - Au'_2(M - x_1), \quad (3)$$

$$f''(x_1) = u''_1(x_1)B - u'_1(x_1)u'_2(M - x_1) + u''_2(M - x_1)A - u'_1(x_1)u'_2(M - x_1), \quad (4)$$

where  $A$  and  $B$  are defined as in P4. Suppose that  $f'(x_1) = 0$ . By (3) and  $B > 0$  (recall P2), we have

$$u'_1(x_1) = \frac{A}{B}u'_2(x_2).$$

By the same calculation as in the proof of P4,

$$u'_1(x_1)u'_2(M - x_1) - u''_2(M - x_1)A = A\left(\frac{(u'_2(x_2))^2}{B} - u''_2(x_2)\right) > 0. \quad (5)$$

By switching the roles of bargainers 1 and 2 and applying the parallel argument, we have

$$u'_1(x_1)u'_2(M - x_1) - u''_1(x_1)B > 0. \quad (6)$$

By (5) and (6), we conclude that the value of (4) is negative.  $\square$

**Claim 2.** There exists at most one  $x_1^* \in (v_1, M)$  such that  $f'(x_1^*) = 0$ .

*Proof.* Suppose by way of contradiction that there exist  $x_1^*, y_1^* \in (v_1, M)$  such that

$$x_1^* < y_1^*, \quad f'(x_1^*) = 0, \quad f'(y_1^*) = 0.$$

Together with Claim 1, there exists  $\varepsilon \in (0, 1)$  such that

$$f'(x_1^* + \varepsilon) < 0, \quad f'(y_1^* - \varepsilon) > 0, \quad x_1^* + \varepsilon < y_1^* - \varepsilon.$$

We define

$$S \equiv \{x_1 \in (v_1, M) : f'(x_1) > 0\} \cap (x_1^* + \varepsilon, y_1^* - \varepsilon).$$

By continuity of  $f'(\cdot)$  (which follows from twice differentiability of utility functions) and  $f'(y_1^* - \varepsilon) > 0$ , we have  $S \neq \emptyset$ .

Since  $S$  is bounded, it has an infimum  $s_1^*$ . By the definition of infimum, there exists a sequence  $\{s_1^k\}_{k=1}^\infty \subseteq S$  such that  $s_1^k \rightarrow s_1^*$ . By the definition of  $S$ ,

$$f'(s_1^k) > 0 \text{ for all } k = 1, 2, \dots.$$

Together with continuity of  $f'(\cdot)$ , we obtain

$$\lim_{k \rightarrow \infty} f'(s_1^k) = f'(s_1^*) \geq 0. \quad (7)$$

By  $f'(x_1^* + \varepsilon) < 0$  and  $s_1^* \geq x_1^* + \varepsilon$ , we obtain

$$s_1^* > x_1^* + \varepsilon. \quad (8)$$

If  $f'(s_1^*) > 0$ , then by continuity of  $f'(\cdot)$  and (8), there exists  $\varepsilon' > 0$  such that

$$f'(s_1^* - \varepsilon') > 0 \text{ and } s_1^* - \varepsilon' > x_1^* + \varepsilon,$$

which implies  $s_1^* - \varepsilon' \in S$ . We obtain a contradiction to the fact that  $s_1^*$  is an infimum of  $S$ . It follows that  $f'(s_1^*) \leq 0$ . Together with (7),

$$f'(s_1^*) = 0.$$

By Claim 1 and (8), there exists  $\varepsilon'' > 0$  such that

$$f'(s_1^* - \varepsilon'') > 0, \quad s_1^* - \varepsilon'' > x_1^* + \varepsilon,$$

which implies  $s_1^* - \varepsilon'' \in S$ . We obtain a contradiction to the fact that  $s_1^*$  is an infimum of  $S$ .  $\square$

We resume the proof of P5. Since the Nash solution is defined by the maximum of the Nash product and it is an interior solution (recall P2), if  $(x_1^*, x_2^*)$  is a Nash allocation, we must have  $f'(x_1^*) = 0$ . Now uniqueness follows from Claim 2.  $\square$

### Proof of P6

*Proof.* The identity equation of P3 is rewritten as  $u'_1(x_1^*)B = u'_2(x_2^*)A$ . The total differential of both sides yields

$$u''_1(x_1^*)dx_1^*B + u'_1(x_1^*)u'_2(x_2^*)dx_2^* = u''_2(x_2^*)dx_2^*A + u'_2(x_2^*)u'_1(x_1^*)dx_1^*$$

Rearranging this equation, we obtain (1). Thus, the second statement follows. The first statement follows from P4 because P4 means that both the denominator and numerator of the right-hand side of (1) are positive.  $\square$

## Appendix 2: Properties of IARA and DARA component functions

In this appendix, we provide examples of the concave IARA and DARA functions, and then, discuss the properties of these functions.

It is worth emphasizing that there exists an IARA function that satisfies the concavity. For example, it is easily checked that the quadratic utility function  $\Psi(x) = x - \alpha x^2, 0 \leq x \leq \frac{1}{2\alpha}$  with  $\alpha > 0$  is such function. Moreover, a logit function like  $\Psi(x) = A(\frac{\exp(x)}{1+\exp(x)} - 0.5)$  with  $A > 0$  is also concave and IARA function.

To consider a class of IARA functions, the analysis of the log-concave probability density function by Bagnoli and Bergstrom (2005) is helpful. They showed that the probability density functions of several kinds of famous probability distribution like normal distribution, logit distribution, etc. satisfy the log-concavity. Since the IARA of  $u$  is equivalent to the log-concavity of  $u'$ , the cumulative density function whose density function satisfies the log-concavity is IARA. Moreover, the cumulative density function of a probability distribution having a continuous density is an increasing function and satisfies concavity in a domain wherein the density is decreasing. Let  $\Phi(\cdot|0, \sigma)$  denote the cumulative density function of a normal distribution with mean 0 and standard deviation  $\sigma$ . Then,  $\Psi(x) = A(\Phi(x|0, \sigma) - 0.5)$  with  $A > 0$  for any  $x \geq 0$  is concave and IARA function.<sup>19</sup> Similarly, we can construct a concave and IARA function from other type of probability distributions.

On the other hand, the concave and DARA utility function is familiar in the literature. For example, as an application of their cumulative prospect theory, Tversky and Kahneman (1992) adopt the DARA utility function like the power-type utility (value) function as follows:  $\Psi(x) = x^\alpha$  with  $0 < \alpha < 1$ . In addition, Saha (1993) considers the expo-utility function defined by  $\Psi(x) = A - \exp(-\beta x^\alpha)$  with some parameter restriction ( $A > 0, \alpha\beta > 0$ ). It is shown that this is concave and DARA function if  $\alpha < 1$ . The one-parameter expo-utility function of Abdellaoui et al. (2007) is defined by  $\Psi(x) = -\exp(-x^\alpha/\alpha)$  for  $\alpha \neq 0$  and  $\Psi(x) = -1/x$  for  $\alpha = 0$ . This is also concave and DARA if  $0 \leq \alpha \leq 1$ . These functions have the common property that if they are concave, they must be DARA (or there is only a narrow region where they can be IARA).

In a utility function defined by (2),  $\lambda$  and the scale parameter of  $\Psi$  determine the shape of the utility function, and the entitlement  $E_i$  becomes the location parameter (see Figure 2). Both models satisfy the following conditions that are emphasized in the prospect theory:

- C1.** diminishing sensitivity
- C2.** loss aversion
- C3.** concave in gain domain and convex in loss domain

However, they are very different in the change in risk attitude with the change in  $x$ . When a component function  $\Psi$  is IARA,  $u$  defined in (2) is IARA in gain and loss domains, too. In addition, since  $\lim_{x_i \rightarrow E-0} r_i(x_i) \leq 0 \leq \lim_{x_i \rightarrow E_i+0} r_i(x_i)$  hold true in this case,  $u$  is IARA in the whole domain. Moreover, if  $\Psi''(0) = 0$ , the Arrow Pratt coefficient becomes a continuous function (see Footnote 19).

<sup>19</sup> This can be checked analytically. By the definition of the density of a normal distribution,

$$\Psi'(x) = A \exp(-\frac{x^2}{2\sigma^2}), \quad \text{and} \quad \Psi''(x) = -A \frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2}).$$

Thus,  $\Psi$  is concave when  $x \geq 0$  and convex when  $x < 0$ . In addition, the Arrow-Pratt coefficient is

$$r_\Psi(x) = \frac{x}{\sigma^2},$$

which is apparently increasing in  $x$ .

In contrast, if  $\Psi$  is DARA, the utility function is DARA in both gain and loss domains. However, we can easily confirm that C3 under the DARA assumption imply that Arrow Pratt coefficient is dis-continuous at the reference point. In fact,  $\lim_{x_i \rightarrow E_i+0} r_i(x_i) = +\infty$  and  $\lim_{x_i \rightarrow E_i-0} r_i(x_i) = -\infty$  hold when the utility function is DARA of Tversky and Kahneman (1992).<sup>20</sup> Therefore,  $u$  under the DARA component function is not DARA in the whole domain.

Mathematically and analytically, there is another significant advantage of using the class of IARA component functions.

**Lemma 2.** If  $\Psi$  satisfies IARA or CARA,  $u_i$  defined in equation (2) with adding some non-negative constant  $A$  of  $A \leq \lambda\Psi(E_i)$  satisfies A2 (without at the reference point).

*Proof.* The proof of the theorem is some modification of the proof of Theorem 1 of Bagnoli and Bergstrom (2005), who investigate the log-concavity of probability density functions. Let  $u$  be the function in equation (2) with adding some  $A$  of  $0 \leq A \leq \lambda\Psi(E)$ . Suppose that  $\Psi$  satisfies IARA, which immediately implies that  $u$  is also IARA because adding  $A$  does not change the property of the first and higher derivatives. Since IARA is equivalent to the concavity of  $\log u'$  and CARA is its linearity, under the assumption of this lemma, we must have

$$(\log u')'' = \frac{d}{dx} \left( \frac{u''(x)}{u'(x)} \right) \leq 0.$$

<!-- In the case of CARA, this becomes the equality. --> <!-- Thus, under the assumption of this lemma, this holds with weak inequality. -->

For any  $x \neq E$ , we can find some  $a \leq x$  such that  $u''(x)u(a)$  is non-positive because when  $x < E$ ,  $u''(x) > 0$  and  $u(0) \leq 0$  by the assumption on  $A$ , and when  $x > E$ ,  $u''(x) < 0$  and  $u(a) > 0$  for any  $a$  with  $E < a < x$ . By using this  $a$ , it is possible to transform  $u''u/u'$  as follows.

$$\frac{u''(x)}{u'(x)}u(x) = \frac{u''(x)}{u'(x)} \int_a^x u'(t)dt + \frac{u''(x)}{u'(x)}u(a) \leq \int_a^x \frac{u''(t)}{u'(t)}u'(t)dt + \frac{u''(x)}{u'(x)}u(a) = u'(x) - u'(a) + \frac{u''(x)}{u'(x)}u(a)$$

where the first equality comes from the fundamental theorem of calculus and the second inequality is from the fact that  $\frac{u''(x)}{u'(x)}$  is non-increasing. Since  $u'(a) > 0$  and  $u''(x)u(a)/u'(x) \leq 0$ , we have

$$\frac{u''(x)}{u'(x)}u(x) \leq u'(x) - u'(a) + \frac{u''(x)}{u'(x)}u(a) \Rightarrow \frac{u''(x)}{u'(x)}u(x) < u'(x) \iff u''(x)u(x) < (u'(x))^2.$$

This is A2. □

Therefore, we can apply our setup and results developed in Sections 2 and 3 by using IARA component functions.

## References

Abdellaoui, M., Barrios, C., and Wakker, P. P. (2007). Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *Journal of Econometrics*, 138(1):356–378.

---

<sup>20</sup>For example, when  $\Psi(x) = x^\alpha$ , the Arrow-Pratt coefficient is

$$r_\Psi(x) = \frac{1 - \alpha}{x - E}.$$

Obviously, it is dis-continuous at  $x = E$ .

- Anbarci, N. and Feltovich, N. (2013). How sensitive are bargaining outcomes to changes in disagreement payoffs? *Experimental Economics*, 16(4):560–596.
- Anbarci, N. and Feltovich, N. (2018). How fully do people exploit their bargaining position? the effects of bargaining institution and the 50–50 norm. *Journal of Economic Behavior & Organization*, 145:320–334.
- Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.
- Aumann, R. J. and Kurz, M. (1977). Power and taxes. *Econometrica*, 45(5):1137–1161.
- Aumann, R. J. and Maschler, M. (1985). Game theoretic analysis of a bankruptcy problem from the talmud. *Journal of Economic Theory*, 36(2):195–213.
- Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic theory*, 26(2):445–469.
- Baranski, A. (2016). Voluntary contributions and collective redistribution. *American Economic Journal: Microeconomics*, 8(4):149–73.
- Baranski, A. (2019). Endogenous claims and collective production: an experimental study on the timing of profit-sharing negotiations and production. *Experimental Economics*, 22(4):857–884.
- Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pages 176–188.
- Binmore, K., Samuelson, L., and Young, P. (2003). Equilibrium selection in bargaining models. *Games and Economic Behavior*, 45(2):296–328.
- Birkeland, S. and Tungodden, B. (2014). Fairness motivation in bargaining: a matter of principle. *Theory and Decision*, 77(1):125–151.
- Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *The American Economic Review*, 90(1):166–193.
- Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827.
- Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2011). The importance of moral reflection and self-reported data in a dictator game with production. *Social Choice and Welfare*, 36(1):105–120.
- Dagan, N. and Volij, O. (1993). The bankruptcy problem: a cooperative bargaining approach. *Mathematical Social Sciences*, 26(3):287–297.
- Ellingsen, T. (1997). The evolution of bargaining behavior. *The Quarterly Journal of Economics*, 112(2):581–602.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Gächter, S. and Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, 51(2):249–263.
- Hennig-Schmidt, H., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2018). Asymmetric outside options in ultimatum bargaining: a systematic analysis. *International Journal of Game Theory*, 47(1):301–329.
- Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic behavior*, 7(3):346–380.

- Hoffman, E. and Spitzer, M. L. (1982). The coase theorem: Some experimental tests. *The Journal of Law and Economics*, 25(1):73–98.
- Hoffman, E. and Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects’ concepts of distributive justice. *The Journal of Legal Studies*, 14(2):259–297.
- Jansen, M. J. and Tijs, S. (1983). Continuity of bargaining solutions. *International Journal of Game Theory*, 12(2):91–105.
- Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, pages 728–741.
- Kamijo, Y. (2023a). A comparative statics on the nash solution. *WINPEC Working Paper Series*.
- Kamijo, Y. (2023b). Fixation of inequality and emergence of the equal split norm: Approach from behavioral bargaining theory. *WINPEC Working Paper Series*.
- Kamijo, Y. (2023c). A note on the risk dominance of the nash demand game. *WINPEC Working Paper Series*.
- Knez, M. J. and Camerer, C. F. (1995). Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behavior*, 10(1):65–94.
- Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.
- Luce, R. D. and Raiffa, H. (1989). *Games and decisions: Introduction and critical survey*. Courier Corporation.
- Luhan, W. J., Poulsen, O., and Roos, M. W. (2019). Money or morality: fairness ideals in unstructured bargaining. *Social Choice and Welfare*, 53(4):655–675.
- Maschler, M., Solan, E., and Shmuel, Z. (2013). *Game Theory*. Cambridge University Press.
- McDonald, I. M. and Solow, R. M. (1981). Wage bargaining and employment. *The American Economic Review*, 71(5):896–908.
- Meyer, H.-D. (1992). Norms and self-interest in ultimatum bargaining: The prince’s prudence. *Journal of Economic Psychology*, 13(2):215–232.
- Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.
- Nydegger, R. V. and Owen, G. (1974). Two-person bargaining: An experimental test of the nash axioms. *International Journal of game theory*, 3(4):239–249.
- O’Neill, B. (1982). A problem of rights arbitration from the talmud. *Mathematical Social Sciences*, 2(4):345–371.
- Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136.
- Raiffa, H. (1982). *The art and science of negotiation*. Harvard University Press.
- Rausser, G. C. and Simon, L. K. (2016). Nash bargaining and risk aversion. *Games and Economic Behavior*, 95:1–9.
- Roth, A. E. (1979). Axiomatic models of bargaining,. In *Lecture Notes in Economics and Mathematical Systems #170*, pages 1–121. Princeton University Press.
- Roth, A. E. (1995). 4. bargaining experiments. In *The Handbook of Experimental Economics*, pages 253–348. Princeton University Press.

- Rubinstein, A., Safra, Z., and Thomson, W. (1992). On the interpretation of the nash bargaining solution and its extension to non-expected utility preferences. *Econometrica*, 60(5):1171–1186.
- Saha, A. (1993). Expo-power utility: A ‘flexible’ form for absolute and relative risk aversion. *American Journal of Agricultural Economics*, 75(4):905–913.
- Svejnar, J. (1986). Bargaining power, fear of disagreement, and wage settlements: Theory and evidence from us industry. *Econometrica*, 54(5):1055–1078.
- Takeuchi, A., Vesteg, R., Kamijo, Y., and Funaki, Y. (2022). Bargaining over a jointly produced pie: The effect of the production function on bargaining outcomes. *Games and Economic Behavior*, 134:169–198.
- Thomson, W. (2003). Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences*, 45(3):249–297.
- Tversky, A. and Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47(2):263–292.
- Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.
- Young, H. P. (1993). An evolutionary model of bargaining. *Journal of Economic Theory*, 59(1):145–168.