

*Contra Private Fairness**

Bart J. Wilson
Economic Science Institute
Chapman University
One University Drive
Orange, CA 92866
Phone: (714) 628-2830
Fax: (714) 628-2881
E-mail: *bartwilson@gmail.com*

April 2009

Abstract: This paper attempts to clarify our understanding of the everyday use of *fair* as we apply it to economic behavior. I first examine the decomposition of *fair* into its semantic primitives and discuss implications of recent research which indicates that the word is untranslatable into any other language, i.e., the concept of *fair* is distinctly Anglo. I also make a Wittgensteinian appeal to context and human sociality as an indispensable tether for what we mean by a fair experience and what we epistemologically know about fairness. The principal implication of this is that rules that guide fair behavior are not located in an individual's private utility function but instead reside in the connections that the individual has to his cultural environs.

JEL Classifications: B4, C9

Keywords: fairness, experimental methodology and philosophy, semantics of economics, rhetoric of economics

* I thank Doug Davis, Erik Kimbrough, Praveen Kujal, Laura Razzolini, and Vernon Smith for comments that have improved this paper. Anonymous referees and participants at the 2007 regional ESA meeting, annual SEA meeting, and the Gruter Conference on Law, Behavior, and Biology also provided valuable feedback. I have also benefited from many helpful conversations with Vernon Smith on rhetoric and method in experimental economics.

How do I know that this colour is red?—It would be an answer to say: “I have learnt English”.

L. Wittgenstein, *Philosophical Investigations*, §381

In the beginning there was the deed.

J. W. v. Goethe, *Faust I*, Scene Six

I. Introduction

Ordinary people appear to use the word *fair* somewhat readily, rather frequently, and quite effortlessly in everyday conversation, and yet, our understanding and use of the term in economics can best be described as muddled. This humbling observation is nowhere more evident than in classroom discussions on the observed behavior in ultimatum games (Güth, Schmittberger, and Schwarze, 1982). In a standard version of the game, proposer a offers to divide a \$10 pie between responder b and himself: $10 - z$ for a and z for b .^{1,2} Responder b can either accept the proposal and both individuals earn their respective amounts of cash in the proposal, or b can reject the proposal sending both individuals home with nothing. If asked why a proposed the modal offer of $z = 5$ to b , a frequently responds with, “Because that’s fair.” If asked why b rejected $z = 2$ from a , b ’s swift response is, “Because that’s not fair.” To both participants, these answers more than adequately convey their intended meaning to the questioner following the arguably artificial exercise they just completed. Yet, when we attempt to construct our understanding of such behavior, we’re addled. An unexamined thought process simultaneously posits that the fairness of a single event, the even split in the ultimatum game, is both the cause for the action and the effect of it:

“Why did proposer a offer $z = 5$ to b ?”

“Because that’s fair.”

“What is a fair outcome in the ultimatum game?”

“Proposer a offers $z = 5$ to b .”³

¹ For various results in ultimatum games see, e.g., Roth (1995), Forsythe, Horowitz, Savin, and Sefton (1994), and Hoffman, McCabe, Shachat, and Smith (1994).

² I will use the ultimatum game as the common example throughout the paper to elucidate the meaning of *fair* in economics, but the implications derived herein are not limited to this game and would apply with the same force to other experimental games that invoke “fairness”. An appendix considers the dictator game.

³ I am indebted to Vernon Smith for pointing out this circularity in similar questions. Nietzsche, in his essay, *On Truth and Lying in the Non-moral Sense*, raises a similar point about “honesty” and then pertinently describes how the many dissimilarities in honest actions are lumped together via the same concept:

We call a man honest; we ask, ‘Why did he act so honestly today?’ Our answer is usually: ‘Because of his honesty. Honesty!—yet again, this means that the leaf is the cause of the leaves. We have no knowledge of an essential quality which might be called honesty, but we do know of numerous individualized and hence

Thus, it appears, as Fehr, Kirchsteiger, and Riedl (1993) succinctly put it in their opening sentence, that “fairness is an elusive term” (p. 437). Unfortunately our current understanding of *fair* is not merely elusive, it is circular, and moreover admitting this to ourselves does not disculpate us for not trying to pin down the semantics of *fair* in economic behavior.

This paper attempts to clarify our understanding of the everyday use of *fair* as we apply it to economic behavior. Is *fair* a lazy label in ordinary discourse that connotes justice or equity or impartiality? To what extent can we incorporate notions of what is fair into our models of utility maximization? What role does human sociality play in an experience deemed to be fair? What is the connection between a fair experience, or the private feelings of what is fair, and the public social interaction that produces such feelings? What is the epistemological relationship between the rules that guide fair behavior and the experience itself that is deemed fair?

Drawing upon the work of Wierzbicka (1996, 2006) I first discuss the decomposition of *fair* into its semantic primitives. Semantic primitives are simple expressions that are indefinable in themselves and which are presumed to be innate human concepts common to all languages. One interesting clue about the content of what is meant by *fair* is that Fletcher (1996) and Wierzbicka (2006) report that the word is one-to-one untranslatable into any other language, i.e., the concept of *fair* is distinctly Anglo in origin. Other languages either directly import the English word, as in the German exclamation “Das ist nicht fair!”; fail altogether to have a comparable word, as is the case for French; or use a single word depending upon the context to cover the separate meanings of *just* and *fair* (e.g., *justo* in Spanish). These observations indicate that the semantic concept of *fair* is culturally portable but not necessarily ubiquitous.

Wierzbicka’s decomposition of *fair* into its constituent semantic primes brings to the foreground many subtle and perhaps hidden nuances of meaning that we rely on when we spontaneously use *fair* in ordinary discourse. First, her definition illuminates in very general terms the nontrivial differences between the meanings of *justice*, *equity*, and *fairness*, something that economic formalizations of fairness conspicuously gloss over. The semantic components also distinguish the private portions of the fairness experience from the public substratum that supports the meaning of *fair*. The principal implication of this is that the meaning of *fair* involves private judgments that emanate from the sociality of the individual. Fairness is neither

non-equivalent actions which we equate with each other by omitting what is unlike, and which we now designate as honest actions; finally we formulate from them a *qualitas occulta* with the name ‘honesty’ [p. 145].

wholly intrinsic nor purely private; rather it is extrinsic and personal. By describing fairness as personal, I mean that it is both subjectively private and objectively public (Polanyi, 1958). This explication of fairness stands in contrast to the traditional approach of explaining fair behavior in a public interaction as emanating from so-called social preferences in a private utility function.

Questioning the utility maximizing approach to explaining fairness may at first jar on the ears of an economist, but we cannot ignore that within the logic of utility maximization the explanandum of *fair* is a perfunctory part of the explanans, i.e., a fair offer of $z = 5$ is part of a social preference for a fair outcome.⁴ This circularity, however, vanishes like Hume's phantom of the night on the appearance of morning when we consider the (later) philosophical insights of Ludwig Wittgenstein. In *Philosophical Investigations (PI)*, published after his death in 1953, Wittgenstein develops the concept of the language game as a tool to guide our understanding of meaning in regularized human activity. A language game is more than just the strategies or actions that an individual can take in an interactive and interdependent setting, but also the mental predicates of feeling, knowing, thinking, and wanting that are interwoven into the action via its concomitant language.

This notion of a language game applies equally to the actions of an ultimatum game's proposer and responder as well as to the theorist and experimenter as we attempt to construct our understanding of the actions of a proposer and a responder. This latter point is important, for with the tool of utility maximization "one thinks that one is tracing the outline of the thing's nature over and over again, and one is merely tracing round the frame through which we look at it. A *picture* [holds] us captive. And we [cannot] get outside it, for it [lies] in our language and language [seems] to repeat it to us inexorably" (Wittgenstein, 1953; §114-5, his emphasis).

Wittgenstein also shows that it is the multiplicity of meaning in our language that creates a fictional doer, fairness, as a hypokeimenon behind the doing and the effecting of fair actions, when instead it is the social practice surrounding the deed that contains the meaning of *fair*. The rule that guides fair behavior is not located in an individual's private utility function but instead resides in the connections that the individual has to his cultural environs.

⁴ See, e.g., Bolton and Ockenfels (2000), Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Fehr and Schmidt (1999). Sobel (2005) contains an au fait summary of work built on this paradigm.

II. The Semantic Prime Decomposition of *Fair*

The study of semantic primes is an approach to understanding meaning in language that involves searching for atomic expressions that presumably make communication across languages possible. Goddard and Wierzbicka (2002) summarize the current list of 61 semantic primitives that have been meticulously examined. They are present in nine diverse languages that span the globe: Polish, Mandarin, Malay, Lao, Spanish, Korean, Mbula, Cree, and Yankunytjatjara. Wierzbicka (1996) succinctly summarizes the goal of this research as an attempt “to pin down the elusive and culture-specific configurations of elements encapsulated in everyday concepts, and to face the formidable complexity of meanings which ordinary people appear to juggle effortlessly in everyday discourse” (p. 233).

One benefit of decomposing a word into its semantic primes is that it serves as a means for elucidating the unconscious nuances of meanings in ordinary conversations. Given the circular nature with which we use *fair* and *fairness* in economics, such an analysis seems appropriate. Table 1 summarizes the subset of semantic primes in a decomposition of *fair*. One indication that a concept is a semantic prime is the circularity with which dictionaries define a term. In her compendium of primes, Wierzbicka (1996) reports as an example that Longman’s Dictionary of the English Language defines *know* in terms of *cognition* and vice versa:

know: to have direct cognition

cognition: the act or process of knowing that involves the processing of sensory information and includes perception awareness, and judgment.

The American Heritage Dictionary of the English Language likewise circularly defines the mental predicate of *think* via *thought* and *cogitation*:

to think: to have a thought;

thought: the act or process of thinking; cogitation;

cogitation: 1. thoughtful consideration; 2. a serious thought.

Circular definitions, however, are not sufficient evidence.⁵ Cross-linguistic comparisons and language acquisition play a critical role in identifying a semantic prime as an innate conceptual building block of language.

Wierzbicka (2006) makes the following observations on everyday usage of *fair* in English:

⁵ After all, the starting point of this paper is that economists circularly use *fair*. However, circular use is not equivalent to an innate semantic concept around which multiple words are circularly defined.

- (1) *fair* is relational to at least one other person;
- (2) being *fair* involves a potential action that someone wants to do and that could be bad for another person; and
- (3) a *fair* interaction is voluntary and done with at least one person.

The first two observations are fairly straightforward. The third distinguishes *fair* from *just*. When a football referee interacts with the players on the field, his calls are deemed to be ‘fair’ or ‘unfair’ and not ‘just’. The referee is present to cooperatively enhance the game for the fans and players. A judge, however, does not similarly “do things with” the plaintiffs and defendants in his courtroom when he metes out justice. Coaches are ‘fair’ (or ‘unfair’), not ‘just’, in their postgame comments on the referees because they are voluntarily part of the game with the referees.⁶ In contrast, a defendant in a courtroom appears involuntarily before, not with, the judge.

Wierzbicka (2006) also takes care to distinguish *fairness* from *equity*. As she notes, equality is implied in such phrases as “your fair share” and “unfair advantage”. The latent equity in these collocations, however, does not universally transfer to such phrases as “fair comments” or “fair play”. There is a sense of sameness in all uses of *fair* that is not present in *equity*. Fairness, as in “fair comments” and “fair play”, implies a consensus about what other people think are the limits of what someone wants to do when it may also be bad for another. In the phrase “your fair share”, the consensus that the share itself is equal coincides with the same way in which other people would concur with the equality of the share in this interaction. But this coincidence is generally not the case, say with a “fair referee” or a “fair teacher”. A referee is fair when the onlookers, coaches, and players deem him so. The implication is that if the consensus about which other people would think about a situation changes, then what is considered to be *fair* may change. So while there may be some common overlaps in the everyday use of *fairness* and *equity*, there are many ordinary uses for which *fairness* does not fit the constricting mold of *equity*. It seems prudent then not to restrict our usage of *fair* in economics to one that connotes *equity*, particularly when the notion of “fair play” is so germane to economic behavior in general and to game theory specifically.

⁶ When the NBA suspends players for leaving the bench to engage an altercation on the court, the question that sports commentators debate is “Are the suspensions fair?”, not “Are the suspensions just?”, nor “Are the suspensions equitable?”. The debate is whether the league’s ruling is justified to the degree that it was executed and whether the rules regarding such conduct are even applicable to this incident. As it will be discussed later, the concept of *fair* includes an agreement that others will think likewise about these circumstances of time and place.

Wierzbicka (2006) proposes the following semantic definition for the expression, *That's fair*:

- (a) I say: people can know that when this person (*a*) did it (*z*),
a did not do anything bad
- (b) if other people know about it they will say the same
because they all think about some things in the same way
- (c) they think like this:
 - (d) *“when people want to do some things with some other people
they know that they can do some kinds of things*
 - (e) *at the same time they know
that they can't do some other kinds of things*
 - (f) *because if they do things like this,
it will be bad for these other people*
 - (g) *they know that people can think that no one will do things like this”*
- (h) people can know that when this person (*a*) did it (*z*), *a* did not do anything like this
- (i) because of this, other people can not say to *a*: “you can't do things like this”

The first component specifies the speaker's judgment that *a* did not do anything bad by doing *z* and that people can know this. If people cannot know whether *a* had done *z*, then there's nothing to express as being *fair*. People must be able to ascertain the circumstances of time and place for *a* doing *z* to pronounce a judgment of it as *fair*.

Components (b) through (h) explicate the underlying assumptions. The assumption in (b) is that there is a consensus that others will likewise think about these circumstances of time and place as the speaker thinks about the same circumstances. If the speaker thinks no one else would concur with the pronouncement of “That's fair”, will the speaker say “That's fair” in the first place? Expressing “That's fair” is often an appeal to the listener to join the speaker in his or her judgment of the circumstances. Component (c) then specifies the thought process that extends through part (h). Part (d) spells out that within this community there are things that a person can do, and simultaneously in (e) there are things that a person cannot do. In other words, there are informal rules to the game by which people in the community voluntarily interact with each other. Why are such rules unstated? Because we could not function (do what we want to

do) if we are pinpointing by conscious awareness the limits within which we make each and every decision.⁷ Furthermore, notice what limits the kinds of things people cannot do—things that will be bad for these other people with whom *a* wants to do things. This is important because what is *fair* involves a tradeoff in welfare between what *a* wants to do and what can be bad for another. Component (g) indicates that others in the community think there are limits to what *a* can justifiably pursue when it conflicts with someone else. This is the key component for which *equity* diverges from *fairness*, for there is nothing in *fairness* that specifies equality of welfare, just a consensus limit. Part (h) then summarizes that *a* did not go beyond the boundaries of the community rules. In other words, the basis to the meaning of *That's fair* rests on what *a* did not do. As part (i) concludes, because of this communitarian understanding of what *a* did not do, people cannot say to *a*, “you can't do things like this”. Part (i) also importantly differs from *just* in that there is a hierarchical component of authority when someone metes out what is and is not *just*. Fairness is not hierarchical, nor is it derived from power.

Let's now apply this to the \$10 ultimatum game in the laboratory when *a* offered $z = 2$ to *b*, to which *b* says, “That's not fair”. The essential difference with *That's fair* is the negation of components (a), (h), and (i).⁸ When *b* says “That's not fair”, he is saying “People can know that when *a* offered $z = 2$, *a* did something bad. And if my fellow students knew that *a* offered $z = 2$, they would say the same because they would think about offers in this experiment in the same way. Their thoughts might go like this, ‘When students want do economics experiments with other students, they know that the proposers can make offers of $z = 4$ or $z = 5$. At the same time, they know that the proposers wouldn't do some other kinds of things such as make offers of $z = 1$ or $z = 2$, because if they do, it would be bad for the responders. My fellow students also know that people would think that no one would do something like offer $z = 1$ or $z = 2$ '. People know that when *a* offered $z = 2$, *a* did exactly this to me. Because of this, my friends would say to *a*, ‘you can't do things like this’.”

⁷ In developing his concept of personal knowledge, Polanyi (1958) makes the distinction between focal and subsidiary awareness in attending to our day-to-day activities. For example, if a pianist shifts his conscious attention from the whole of a piece of music (focal awareness) to the particular motions of his fingers (subsidiary awareness), he becomes confused and loses the coordination on which he was relying (p. 55).

⁸ Someone does not make an offer $z = 2$ thinking that he or she is being unfair. In this example there is a lack of agreement between *a* and *b* as to what is fair in these circumstances. The semantic decomposition spells out that the concept of fairness is a product of agreement and that unfairness is the lack of agreement. Furthermore, as an agreement, fairness cannot be private (see section IV).

Notice two things about the semantic decomposition. First, the mental predicate “know” pervades the definition. People know what *a* did, know what other people will think about what *a* did, know what kinds of things *a* can do that are not bad to *b*, know what kinds of things *a* can do to *b* that are bad, know that people can think that no one will do things to *b* that are bad, and know because of all this knowledge that other people would say to *a* that “you can’t do things like this”. But whence all this knowledge on the part of *b*? And where is it captured in a utility maximization problem?

The second observation to note is that the explanans is not included in the explanandum of *fair*. This definition composed of semantic primes avoids circularity by appealing to two pieces of knowledge external to *b*—(1) what other people know about what others would think and do and (2) the limits thereof for actions that are bad. This knowledge of the external social world may be applied internally as part of what *b* knows when he deems the offer of $z = 2$ to be unfair, but the source is not an inner parameter of a utility function. What makes *b*’s fellow students think that proposers can make offers of $z = 4$ or $z = 5$ but that offers of $z = 1$ or $z = 2$ are bad? Such offers are not frozen in the mists of time and place as either good or bad. The next section makes a Wittgensteinian appeal to context and human sociality as an indispensable tether for what we mean by a fair experience and what we know about fairness. But what is the connection between the fair experience, i.e., the private, personal thoughts of what is fair, and the public social interactions that produce such thoughts? Can *a* know what *b* will privately deem to be fair? If so, then how?

III. A Fair Experience, Sociality, and Knowledge

The goal of this section is to explicate the role that sociality plays in a fair experience vis-à-vis knowledge without taking either of two false steps: contending that specific social interactions are a necessary condition for a fair experience and dismissing sociality as not mattering at all. What we are looking for is the interaction of sociality with what one personally experiences as fair. The first step is to understand the role that knowledge plays in this interaction.

Wittgenstein takes this question head on in his well-known arguments against private language. The key issue is knowledge and what one is relying on to articulate what is fair when one experiences it. The first point is simply that one need not personally encounter an

experience as fair to know what fair is, and if one admits the possibility of knowing without a specific experiential tether, then it becomes easier to understand how sociality conditions a fair experience. Wittgenstein succinctly makes this argument in §293 in *Philosophical Investigations*:

If I say to myself that it is only from my own case that I know what the word “pain” means—must I not say the same of other people too? And how can I generalize the *one* case so irresponsibly?

Now someone tells me that *he* knows what pain is only from his own case!—Suppose everyone had a box with something in it: we call it a “beetle”. No one can look into anyone else’s box, and everyone says he knows what a beetle is only by looking at *his* beetle.—Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.—But suppose the word “beetle” had a use in these people’s language?—If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: for the box might even be empty.—No, one can ‘divide through’ by the thing in the box; it cancels out, whatever it is (his emphasis).

Notice how Wittgenstein makes his point with this thought experiment. First, he takes this amorphous concept of a mind and puts it in a confined box. It may be formless inside the box but it is still contained within a physical body. Moreover, the situation involves a group of people with their own boxes, and we are led to imagine a group of distinct, physically-contained minds.

Inside the box is a beetle, say a fairness beetle. Given our innate curiosity with other people’s things, we have an urge to look inside other people’s boxes, but we can’t—the fairness beetle is personal. No one can look into anyone else’s mind to know what fair is. When *a* makes an offer of *z* to *b*, *b* is the one who personally experiences *z* as fair or unfair. What Wittgenstein says of patience can be said of fairness: “One plays patience by oneself” (*PI*, §248).

In the course of a group discussion, everyone says they know what a fairness beetle is from inspecting the contents of their own box, even though *a*’s fairness beetle is red, *b*’s is blue and *c*’s isn’t even a beetle (it’s a spider). But everyone can only look at their own fairness beetle at this time and place, for “one might even imagine such a thing constantly changing”.

In the context of a discussion, say on a particular ultimatum game, each individual uses *fair* in a comprehensible way to further the discussion. When called upon to describe what they are talking about, without conscious awareness each person uses *fair* for what he means. And even if all of their boxes are empty, that is, they have never experienced what fairness is, everyone in the group still shares an understanding of the meaning of fairness by virtue of their ability to coherently discuss it in context of an ultimatum game.⁹ This seems rather apposite for our inquiry of the word *fair*, which, as Wierzbicka shows, is distinctly Anglo in origin and yet directly portable into other languages.

So, if each mind has its own conception of what fair is, how can everyone outwardly discuss it? Because as part of their shared sociality in regularized human activity, they have a learnt system of informal and tacit rules on how people interact. Each individual understands what makes people act because sociality is a public system, a public system of knowing what kinds of things *a* can do that are not bad to *b*, knowing what kinds of things *a* can do that are bad to *b*, and knowing that people can think that no one will do bad things to *b*. Each individual action fits within a pattern of informal and tacit rules by which we interact, just as words fit into a sentence within a set of rules that we never learn by rote (Pinker, 1994). Understanding these rules, implicitly and explicitly, means that one understands human sociality. As Wittgenstein puts it, “To understand a sentence means to understand a language. To understand a language means to be master of a technique” (*PI*, §199). Furthermore, it takes time to master human sociality: “The child, I should like to say, learns to react in such-and-such a way; and in so reacting it doesn’t so far know anything. Knowing only begins at a later level” (*On Certainty*, §538).

In sum, personal experiences with fairness cannot be separated from sociality but can be modulated by it. If individuals do not have experiences that a community deems to be fair, then almost reflexively, it may be possible to dispense with the need to consider fairness altogether within regularized activities. The notable lack of a comparable word for *fair* in any other language points rather markedly to this as a possibility. So, in this sense a particular situation that could be labeled as *fair* is necessary for the social use of *fair* (for how else could we talk about it?), but having a fair experience is not equivalent to recognizing and deeming an

⁹ “It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say” (*PI*, §142).

experience as fair. We learn to recognize and deem experiences as fair by learning to be social; we do not learn about what is fair via a particular experience (or maximizing a utility function).¹⁰ In terms of the ultimatum game, the experience of getting an offer of $z = 5$ does not constitute b 's learning that the offer is fair. Rather, from having been a part of regular human interaction, b identifies the offer of $z = 5$ as fair. Maximizing a utility function with interdependent preferences on the outcomes treats having a fair experience as one and the same with deeming an experience as fair. As the parable of the beetle makes clear, understanding an experience as fair is an occurrent state of a person *only within a social practice governed by tacit rules*.

IV. Rules of Fairness and the Fair Experience

The next natural question to ask is “What is the epistemological relationship between a tacit rule that guides fair behavior and its application in deeming an experience as fair?” Let's first consider the following case:

[A] pupil has mastered the series of natural numbers. Next we teach him to write down other series of cardinal numbers and get him to point to writing down series of the form

$0, n, 2n, 3n, \text{etc.}$

at an order of the form “ $+n$ ”; so at the order “ $+1$ ” he writes down the series of natural numbers.—Let us suppose we have done exercises and given him tests up to 1000.

Now we get the pupil to continue a series (say $+2$) beyond 1000—and he writes 1000, 1004, 1008, 1012.

We say to him: “Look what you've done!”—He doesn't understand. We say: “You were meant to add *two*: look how you began the series!”—He answers: “Yes, isn't it right? I thought that was how I was *meant* to do it.”—Or suppose he pointed to the series and said: “But I went on in the same way.”—It would be no use to say: “But can't you see....?”—and repeat the old examples and explanations.—In such a case we might say, perhaps: It comes natural to this person to understand our order with our explanations as *we* should understand the order: “Add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on.”

Such a case would present similarities with one in which a person naturally reacted to the gesture of pointing with the hand by looking in the direction of the

¹⁰ Moreover, this “learning”—like ‘culture,’ ‘rationality,’ and ‘intelligence’—is not an explanation for anything, but rather a phenomenon that itself requires explanation” (Tooby and Cosmides, 1992, p. 122). That explanation, however, is beyond the scope of this paper.

line from finger-tip to wrist, not from the wrist to finger-tip (*PI*, §185; his emphasis).

Wittgenstein is using an irregular case of the relationship between a rule and its application to illustrate the regular case. The aim is to demonstrate the principle that there is nothing in our training in social processes that forces a particular application of a rule. The same schooling and the same directive brings something very different to the mind of Wittgenstein's pupil than to us (we've all had a student like his). So what makes the pupil's application incorrect and ours correct? The rule receives its meaning not from rule itself and the specific order to execute it, but from the entire practice of using natural numbers in mathematics. Without any indecisiveness, our training in the rule brings before our mind how to act: 1000, 1002, 1004, 1006, etc. Where then is the conflict? *Not in the pupil's response and the application of the rule, but in the pupil's response and our response.* The pupil is applying another rule before his mind to this particular case (say, for $n \in \mathbb{N}$, +2 for $n \leq 1000$, +4 for $1000 < n \leq 2000$, +6 for $2000 < n \leq 3000$, etc.) and simply repeating the "the old examples and explanations" is not going to resolve the conflict between our expected response and the pupil's actual response. (Recall that the pupil has only been tested up to 1000.) The source for the different responses cannot be found in the rule itself, though we are disposed to look there, but rather in the entire practice of teaching natural numbers, assigning exercises, and grading the responses. We have an expectation about the practice of natural numbers, not the specific rule involving them, that Wittgenstein's pupil personally does not share.

Now let's apply this argument to the \$10 ultimatum game.¹¹ If a applies a rule of fairness to the ultimatum game and offers $z = 4$ to b and b shares the expectation about the practice surrounding this experimental exercise, then b accepts the offer.¹² However, if a instead applies a rule of fairness that prescribes an offer of $z = 2$ to b and b does not share this expectation about the entire practice surrounding this task, then b will reject the offer. Notice that it makes no difference that Wittgenstein's pupil is given an explicit mathematical rule and that a 's learnt rule of fairness is tacit. The argument is the same. The source for the different responses of a and b cannot be found in the rule itself because the entire custom of social interactions is what

¹¹ An appendix applies the semantic prime and language game analyses to the dictator game.

¹² Camerer (2003) reports that the modal offer is regularly 40-50 percent of the pie in a typical ultimatum game of the Forsythe et al. (1994) variety.

matters.¹³ If there is conflict (or a lack of agreement) concerning $z = 2$, it is observed in the actions of a and b , not in a 's rule of fairness and the action of b .

Furthermore, no matter how inclined we are, we cannot ask the question "How can a rule of fairness show us what a and b will do in this ultimatum game?" That's asking us to look in other people's boxes to sneak a peak of their fairness beetle. Wittgenstein instead proposes a different place to look:

Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?—Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.

[Interlocutor:] But that is only to give a causal connexion; to tell how it has come about that we now go by the sign-post; not what this going-by-the-sign really consists in.

[Wittgenstein:] On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom (*PI*, §198).

A sign-post conveys meaning to a traveler and is not simply a piece of metal sticking out of the ground because sign-posts are a regular part of the social practice of traveling. The interlocutor misses the point that the connection is between the traveler's action and the regular use of sign-posts, not between the action and some interpretation of the piece of metal as a rule to follow. Sociality imparts the meaning of the sign-post. If someone is completely foreign to the community and hence unfamiliar with sign-posts, then what a sign-post is to us is to him but a piece of metal without any meaning.

What does a rule of fairness have to do with a 's actions? Proposer a has been trained through his sociality to react to a rule of fairness in a particular way and so offers $z = 5$. An objection is that this answer fails to spell out what following the rule of fairness really consists of, say in a utility function. But that objection misses the point that a follows a rule of fairness only *insofar as there exists regular social use of a rule of fairness*, a custom. The meaning of fairness emanates from the sociality that surrounds what a and b do. If a is unfamiliar with the

¹³ When proposers (sellers) are primed in the instructions of an ultimatum game to "consider what choice you expect the buyer [responder] to make [and to] also consider what you think the buyer expects you to choose," Hoffman, McCabe, and Smith (2000) find that offers to the responders significantly increase. In other words, more consideration of the social custom leads proposers to focus on the possibility that responders may not agree with a lower offer and hence they make higher offers.

rule of fairness for this particular social interaction in an ultimatum game, z will equal 1, the subgame perfect Nash equilibrium offer.¹⁴

What remains to be discussed are some epistemic considerations of the relationship between a rule that guides fair behavior and its application. When someone points with their index finger, without conscious awareness we attend to the direction of the wrist to the tip of the finger (and not vice versa). That is the rule we have learned and committed to body memory since being a toddler. Each of us has done this many times as part of being social and communicating with one another. What does it mean for us to follow or obey a rule? If we simply just do it without pondering justifications for why we do it, just as we follow the line of sight from a wrist to a pointing finger without conscious awareness, we are obeying a rule. “If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: ‘This is simply what I do’” (*PI*, §217). This is not to say that in an ultimatum game, a gives no thought to his decision in offering $z = 5$ to b , for thought has everything to do with his action in an experiment. Rather, what I mean is that a gives no justification for obeying a rule of fairness in taking the action of $z = 5$.

This brings us full circle to the questions and answers that were posed in the introduction about how we think about fairness in an ultimatum game:

“Why did proposer a offer $z = 5$ to b ?”

“Because that’s fair.”

“What is a fair outcome in the ultimatum game?”

“Proposer a offers $z = 5$ to b ?”

Subjects and experimenters alike do not include any justification for obeying a rule of fairness in an ultimatum game. Why? Because “‘obeying a rule’ is a practice” (*PI*, §202). Within a social framework, following an unstated rule is a practice because it is simply what we do. Wittgenstein also puts it this way: “When I obey a rule, I do not choose. I obey the rule *blindly*” (*PI*, §219; his emphasis). By *blindly*, he does not mean as an automaton devoid of thought, but rather as someone acting with blinders, without reflection upon an entire field of vision. It is telling how easily we observe people using *fair* in everyday conversation. People do not struggle for the “right” word when they use *fair* to describe an experience.¹⁵ Moreover, no reflection is

¹⁴ And the only individuals that I have seen actually make this offer for real stakes are students of economics, undergraduate and graduate. Somehow their training has supplanted the rules of fairness that ordinary people apply.

¹⁵ One easily imagine how unlikely it would be for someone to say, “That’s not...Wait, the word’s on the tip of tongue...What is it?...Yes, that’s not fair!”

necessary for *fair* to be appropriate for the circumstance. But that does not mean that we are not reflecting on the situation itself when use the word *fair*. This is the sense in which we blindly apply a rule of fairness.

When our students effortlessly use the word *fair* to describe their behavior in an ultimatum game, it is because there is a mutual sense of the obvious, with respect to the listener, that a rule of fairness is applicable to this situation. Learning bedrock practices is how we generate a shared sense of the obvious. Hence, as a practice, obeying a rule of fairness is automatic within a framework of regularized social activity.¹⁶ Furthermore, “to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule ‘privately’: otherwise thinking one was obeying a rule would be the same thing as obeying it” (*PI*, §202; his emphasis). One cannot obey a rule involving fairness “privately” because one must share the practice with another for obeying a rule to have meaning. It now becomes obvious why fairness is an elusive term in economics. When the idea of a shared practice or custom is missing from our reconstruction of motives (or “utility”) from observed behavior, fairness becomes the cause and the effect of an action. Sociality is the way out of this circular reasoning. Sociality generates bedrock practices such as obeying a rule of fairness that serves as the grounding by which we understand the meaning of our everyday behavior.

V. Discussion and Conclusion

The spirit of this paper has much in common with Levitt and List (2007) who question how much we can learn by comparing experimental behavior to models of social preferences.¹⁷ The difference between their paper and mine, however, is that they follow the tradition in economic modeling by distilling “moral” rule-following into a private utility maximization problem.¹⁸ While their model may include such variables as a social norm (*n*) and scrutiny (*s*), each agent still follows a rule of fairness purely privately. But a domain-general model of utility maximization does not nor cannot predict when someone applies a rule of fairness to this specific situation. The determiner *this* is the operative word. As Wierzbicka’s semantic decomposition

¹⁶ This sentence is *not* stating that ‘fairness’ is a bedrock concept. Applying or “obeying” a rule of fairness is a bedrock practice.

¹⁷ In addition to the references noted above, see Charness and Rabin (2002), Cox, Friedman, and Gjerstad (2007), Falk and Fischbacher (1999), and Levine (1998).

¹⁸ McCloskey (2006) goes to a considerable effort to show how this framework fails to capture the human virtues of love, faith, and hope.

makes clear, there is a lot of knowledge about community standards and rules that cannot be captured by a utility function. Thus, it is the context of this interaction which determines if a rule of fairness is applicable.¹⁹

While the character of any individual's experience is subjectively private in that only "I" have "my" fair experience (my beetle in the box), the common sociality that surrounds the experience and which allows us to talk about fairness is objectively public. Sociality is manifested in the rules of fairness that people follow in "blindly" applying it to a specific situation. Fairness doesn't "do" anything and everyday people don't maximize utility functions, but people do follow rules as part of regularized human interactions.²⁰ The social practice of following a rule of fairness consists only in "doing this" or "doing that", the deed.

What then is the prescription for economic research on fairness? If the sociality of custom is critical to understanding when people apply a rule of fairness, then the first step is to design experiments that cultivate contrasting autobiographical experiential memories/knowledge in the laboratory. The larger implication is that our experimental designs need to step out from the shadow of game theory, and more specifically, from one-shot games that we optimistically

¹⁹ See Wilson (2008) for a discussion of the inextricable role that context plays in a reciprocating act and its intention.

²⁰ A skeptic might try to reposition the argument by replying that economists employ a framework of social preferences to gain insights "as-if" people solve a utility maximization problem, as some sort of proximate, rank ordering mechanism. This is similar to a referee's genuine question as to whether social preference models are meant for understanding the consequences of fairness or for explicating the cause of fairness. The problem is that this is a false dichotomy in describing what economists actually do when they "do science". We do both. Moreover, as Wilson (2008) discusses, the scientific method of inquiry can conflate an experimenter's cause for an observed outcome (perhaps motivated by a theorist's mathematical prediction of such behavior) with a subject's reason for acting. When we develop models of social preferences to make new predictions, as Camerer (2003, p. 101) prescribes, we may be tempted through our language to posit social preferences as an agent of cooperation, when it is the individual's action embedded in a pattern of social custom that constitutes the agency. Wilson (2008) argues how social preferences are not the cause of reciprocity, and a similar argument would apply to fairness here. Furthermore, the discussion in section IV further argues the futility of using social preference models for understanding the consequences of fairness. What criteria can distinguish the actual actions taken by individual from the inner state of social preferences that consequently result in fairness? There are none. As with Wittgenstein's pupil, we can never know the rule of fairness that is before the mind's eye of another person; only the expression of a rule of fairness is observable in the form of the actions that people take. Is (private) utility maximization then likewise futile for gaining insights into basic consumer theory? No, because as Afriat (1967) proves, there are external observable criteria in consumer theory that *rationalize* a person's behavior ex post: consumption bundles purchased at alternative prices. Social preference models that rely on utility maximization do not have such external observable criteria to rationalize behavior. Andreoni and Miller (2002), Cox, Friedman, and Sadiraj (2008), and Fisman, Kariv, and Markovits (2007), however, assume just this. Cox et al. (2008) assert that a price p exists for another's payoff in the budget set and the other two experiments artificially induce one in a dictator game. But from observations of everyday behavior, what is the meaning of a p for another's payoff that constrains one's utility maximization of social preferences? A price in consumer theory is a portion of my income at an externally fixed rate that I exchange with another person for goods. In the realm of regularized social interactions, there is no external and hence explicitly observable and semantically meaningful p .

construct as neutral and then naïvely assume to be acted upon just as neutrally. This would involve posing one-shot tasks in terms of specific contextual references with a goal of tracing out the general principles under which people apply rules of fairness. One way to elicit the shared notions of fairness may include asking directly participants how they view decisions they make in different contexts or placing the decisions within in an instant message system that records their unguarded comments as they carry out their experimental tasks.²¹

Because fairness is not fixed outside regularized human activity but rather is semantically interpreted within social practice, economic experiments can use repeated interactions to see where people agree as to what is fair (perhaps also within an instant message system). Many experimental economists eschew implementing repeated games because the Folk theorem demonstrates that nearly any outcome can be supported as a solution (for infinitely repeated games or games with uncertain endings). The Folk theorem, however, merely suffices to describe a plethora of conditions under which people agree to cooperate. It does not explain the process and customs by which people arrive at an agreement. Experimental treatments are a means to elucidate that process. Repeated interactions are an immediate and evolving context which people draw upon to mutually agree on what rules of fairness apply to the current circumstance (probably much more so than any induced external contextual referents that one may implement for a one-shot game). Cooperative outcomes are the product of human agreement on the social context of the interaction, and repeated interaction simply makes it easier to match mutual expectations. The question for experimental research on fairness is why people do or do not share expectations on what is fair in particular circumstances.

The pervasive thought in the field, however, is that one-shot games “insulate behavior from the incentives for cooperation and reciprocity that are present in repeated games”, i.e., one-shot games “allow us to abstract away from issues of learning and attempts to manipulate others’ beliefs, behavior, or preferences” (Goeree and Holt, 2001; pp. 1403 and 1418). As Smith (2008) discusses, the maintained assumption here is that our subjects view their tasks as a game theorist

²¹ An example of the former is Franciosi, Kujal, Michelitsch, Smith, and Deng (1995), who use two different survey questions to vary the context in which a hardware store raises the price of shovels following a snowstorm. (This paper probes below the surface of the survey studies of Kahneman et al. (1986a, 1986b).) When a justification is provided (the higher price prevents a stockout for the regular customers), only 19.5% respond that it is “unacceptable” or “completely unacceptable” for the store to raise its price. Without the justification, 32.4% consider the price increase of “unacceptable” or “completely unacceptable”. Crockett, Smith, and Wilson (2008) and Kimbrough, Smith, and Wilson (2008) use a natural language instant messaging system as part of their experimental institutions to examine social attitudes in the discovery process of exchange and specialization.

views the task, and that is plainly not the case. Experimental participants bring with them tacit knowledge on social practice: they know what other people will think about what people do, know what kinds of things people can do that are not bad to others, know what kinds of things people can do to others that are bad, know that people can think that no one will do things that are bad, and know because of all this knowledge that other people would say that “you can’t do things like this”.²²

In sum, the notion of fairness in the purely private domain of a utility function, severed from the sphere of social practice, is simply incoherent, for the concept of fairness as a rule of behavior is ultimately the concept of a certain form of fairness in custom. A paraphrase of the opening quotation carries a double meaning for economics: How do I know that this outcome is fair?—It would be an answer to say: “I have learnt English”.

References

- Afriat, Sidney N. 1967. “The Construction of a Utility Function from Expenditure Data.” *International Economic Review*, 8: 67-77.
- Andreoni, James and John Miller. 2002. “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.” *Econometrica*, 70: 737-53.
- Bolton, Gary E. and Axel Ockenfels. 2000. “ERC: A Theory of Equity, Reciprocity, and Competition.” *American Economic Review*, 90(1): 166-193.
- Bornstein, Gary and Ilan Yaniv. 1998. “Individual and Group Behavior in the Ultimatum Game: Are Groups More “Rational” Players?” *Experimental Economics*, 1(1): 101-108.
- Brosig, Jeannette, Thomas Riechmann, and Joachim Weimann. 2007. “Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior.” Faculty of Economics and Management, University of Magdeburg.
- Camerer, Colin. 2003. Behavioral Game Theory: Experiments in Strategic Interaction. New York: Sage.
- Cason, Timothy N. and Vai-Lam Mui. 1998. “Social Influence in the Sequential Dictator Game.” *Journal of Mathematical Psychology*, 42: 248-265.

²² As a last-ditch effort to rescue the paradigm, a critic might posit that if “the social part of fairness is already largely completed when humans become grown ups, one could argue that for them fairness can be captured in a private utility function.” Regardless of whether one adopts the social custom of applying a rule of fairness just last week or decades ago in one’s childhood, an individual cannot apply a rule of fairness privately in the present because one must share the practice *hic et nunc* with another for applying the rule to have any meaning in the present.

- Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-869.
- Cherry, Todd L., Peter Frykblom, and Jason F. Shogren. 2002. "Hardnose the Dictator." *American Economic Review*, 92(4): 1218-1221.
- Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59: 17-45.
- Cox, James C., Daniel Friedman, and Vjollca Sadiraj. 2008. "Revealed Altruism." *Econometrica*, 76(1): 31-69.
- Crockett, Sean, Vernon L. Smith, and Bart J. Wilson. 2008. "Exchange and Specialisation as a Discovery Process," *Economic Journal*, forthcoming.
- Dufwenberg, Martin and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47: 268-298.
- Falk, Armin and Urs Fischbacher. 1999. "A Theory of Reciprocity." Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 6.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, 108(2): 437-459.
- Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.
- Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. "Individual Preferences for Giving." *American Economic Review*, 97(5): 1858-1876.
- Fletcher, George. 1996. Basic Concepts of Legal Thought. New York: Oxford University Press.
- Forsythe, Robert, Joel Horowitz, N.E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6(3): 347-369.
- Franciosi, Robert, Praveen Kujal, Roland Michelitsch, Vernon Smith, and Gang Deng. 1995. "Fairness: Effect on Temporary and Equilibrium Prices in Posted-Offer Markets." *Economic Journal*, 105 (431): 938-50.
- Goddard, Cliff and Anna Wierzbicka. 2002. Meaning and Universal Grammar: Theory and Empirical Findings. Volumes 1 and 2. Amsterdam: John Benjamins.
- Goeree, Jacob and Charles Holt. 2001. "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions." *American Economic Review*, 91(5): 1402-1422.

- Güth, Werner, Rolf Schmittberger and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3(4): 367-388.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7(3): 346-380.
- Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 2000. "The Impact of Exchange Context on the Activation of Equity in Ultimatum Games." *Experimental Economics*, 3(1): 5-9.
- Kahneman, Daniel, Knetsch, Jack, and Richard Thaler. 1986a. "Fairness and the Assumptions of Economics." *Journal of Business*, 59(4): S285-300.
- Kahneman, Daniel, Knetsch, Jack, and Richard Thaler. 1986b. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*, 76(4): 728-741.
- Kimbrough, Erik O., Vernon L. Smith, and Bart J. Wilson. 2008. "Historical Property Rights, Sociality, and the Emergence of Impersonal Exchange in Long-distance Trade." *American Economic Review*, 93(3): 1009-39.
- List, John A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy*, 115(3): 482-494.
- Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1(3): 595-622.
- Levitt, Steven D. and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.
- McCloskey, Deirdre. 2006. The Bourgeois Virtues: Ethics for an Age of Commerce. Chicago: University of Chicago Press.
- Nietzsche, Friedrich. 1999. The Birth of Tragedy and Other Writings. Eds. Raymond Geuss and Ronald Speirs. New York: Cambridge University Press.
- Oxoby, Robert J. and John Spraggon. 2008. "Mine and Yours: Property Rights in Dictator Games." *Journal of Economic Behavior and Organization*, 65: 703-18.
- Polanyi, Michael. 1958. Personal Knowledge: Towards a Post-Critical Philosophy. Chicago: University of Chicago Press.
- Pinker, Steven. 1994. The Language Instinct. New York: William Morrow and Company.

- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83: 1281-1302.
- Roth, Alvin. 1995. "Bargaining Experiments." In Handbook of Experimental Economics. Eds. John Kagel and Alvin Roth. Princeton, NJ: Princeton U Press, 253-348.
- Schurter, Karl, and Bart J. Wilson. 2008. "Justice and Fairness in Dictator Games." *Southern Economic Journal*, forthcoming.
- Smith, Vernon L. 2008. Rationality in Economics: Constructivist and Ecological Forms. New York: Cambridge U Press.
- Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, 43(2): 392-436.
- Tooby, John, and Leda Cosmides. 1992. "The Psychological Foundations of Culture," In The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Eds. Jerome H. Barkow, Leda Cosmides, and John Tooby. New York, NY: Oxford University Press, 19-136.
- Wierzbicka, Anna. 2006. English: Meaning and Culture. New York: Oxford University Press, 141-167.
- Wierzbicka, Anna. 1996. Semantics: Primes and Universals. New York: Oxford University Press.
- Wilson, Bart J. 2008. "Language Games of Reciprocity." *Journal of Economic Behavior and Organization*, 68(2): 365-377.
- Wittgenstein, Ludwig. Philosophical Investigations. Translated by G.E.M. Anscombe. Malden, MA: Blackwell Publishing, 1953.

Table 1. Semantic Primes in *Fair*

<i>Substantives</i>	I, PEOPLE, SOMEONE, SOMETHING/THING
<i>Determiner</i>	THIS, THE SAME, OTHER
<i>Quantifiers</i>	ALL
<i>Evaluators</i>	GOOD, BAD
<i>Speech</i>	SAY
<i>Mental predicates</i>	THINK, FEEL, WANT, KNOW
<i>Action</i>	DO
<i>Time</i>	WHEN, AFTER
<i>"Logical" concepts</i>	NOT, CAN, BECAUSE, IF
<i>Taxonomy, paronymy</i>	KIND OF
<i>Similarity</i>	LIKE

Appendix A: Fairness and the Dictator Game

To test whether offers of $z > 1$ in the ultimatum game are attributable to the “fairness hypothesis”, Forsythe et al. (1994) introduce the dictator game in which b no longer has a right to veto the offer of z , i.e., a 's decision is final. Let's now apply the semantic prime and language game analyses to this game when a offers $z > 0$ to b . First, I note that someone does not make a positive offer thinking that he or she is being unfair. The semantic decomposition spells out that the concept of fairness is a product of agreement and that unfairness is the lack of agreement. Unlike the ultimatum game, agreement with b is not a strategic and hence focal consideration in the dictator game. So where would a consideration of what's fair come into play in these circumstances? More to the point, with whom is there agreement? The experimenter?²³ But if a is concerned with agreement with the experimenter, could not a also be concerned about agreement with b even though b has no strategic options? The answer is possibly both of these considerations of agreement are important.

The upshot is that the dictator game is perhaps not as straightforward as it may seem at first blush.²⁴ If the experimenter supplies a with the money in a dictator game, does a apply a rule of fairness to these circumstances? Or possibly another rule of social behavior such as a rule of equity or a rule of just desert? If the experimenter cannot know what decision any specific person makes, does a still apply a rule of fairness in this socially distant circumstance? If the experimenter gives a the opportunity to earn the money to allocate, does a apply a rule of fairness or equity or just desert? Or if the experimenter gives a the right to allocate the money based upon his class standing, does a apply a rule of fairness or equity or just desert?

When Forsythe et al. find that a 's still offer 24% of the pie even though b has no say in the matter, they conclude that offers in the ultimatum game “cannot be fully explained by a taste for fairness” (p. 362). I contend that we may need to take care in coming to such conclusions because subjects may apply tacit rules of social behavior conditional upon the rules of the

²³ Cason and Mui (1998) and Brosig et al. (2007) find that subjects who repeatedly play dictator games on average become more self-regarding over time. A question that might come to mind is, “How does this observation square with the intuition of the ‘consensus about other people think are the limits of what someone wants to do when it may also be bad for another’ (p. 4) given that the ‘consensus’ is unlikely to have changed during this experiment?” My reply would be that it is not self-evident that any general agreement is frozen in time and place for such a peculiar exercise, which is, strictly speaking, not even a “game”. One plausible hypothesis is that an initial agreement with the experimenter is eroding. The question that needs to be asked via another experimental design is why people become more self-regarding over time.

²⁴ The dictator game appears straightforward only from a view that subsidiarily assumes a game theoretic perspective. Fortunately people are not game theoretic automatons.

experiment. Nevertheless, this introduces a puzzle. Why do subjects still offer over 20% of the pie on average? Under protocols that ensure double blind anonymity, Hoffman et al. (1994) find that subjects only offer 7% of the pie in the dictator game. The striking conclusion is that agreement with the experimenter is important when the experimenter knows *a*'s identity and the decision that *a* makes, but not when the experimenter is blind to *a*'s identity. When *a*'s must earn the money to be divided, Cherry et al. (2002) find that only 3% of dictators offer anything with double blind anonymity, and 0% offer anything in Oxoby and Spraggon (2008). Apparently, under these circumstances a rule of fairness goes completely out the window and a rule of just desert replaces it.²⁵ A question that Smith (2008) raises is how many times in our everyday lives do we make decisions to allocate money that we have not earned ourselves? Another question that I would add is why do we apply a rule of fairness to other people's money and not our own earned cash?²⁶ More generally, when do we apply a rule of just desert and when a rule of fairness?

To summarize, because agreement is built into the strategic structure of the ultimatum game, a rule of fairness is focally applicable, but agreement is only subsidiarily a part of the dictator game. This may answer the question as to why offers in the ultimatum game vis-à-vis the dictator game are much more robust with respect to double blind protocols (Bornstein, and Yaniv, 1998). The dictator game is more sensitive to double blind protocols because dictators apply a rule of just desert instead of a rule of fairness when there is no one else with whom to agree (save one's self) and quite dramatically so when they have earned the money.²⁷

²⁵ Schurter and Wilson (2008) argue that earning the right to be the dictator and not the receiver (as opposed to earning the money as the dictator) confounds a rule of just desert (a meritorious ranking) with the rule of fairness in the procedure/opportunity that determines the dictator. They conduct a dictator game in which the external merit of seniority (as measured by the total number of credit hours completed or in progress) determines the role of dictators and that all subjects have explicitly agreed to this criteria before proceeding into the experiment. This distribution of offers is very similar to one in which the dictators are determined by a quiz (which the subjects have also explicitly agreed to before proceeding into the quiz and experiment). The conclusion is that subjects are applying a rule of just desert rather than a rule of fairness in the quiz treatment.

²⁶ A skeptic might note that "fairness behaviour exhibits a huge variance among different individuals" and "wonder how this is compatible with the stated consensus" from section 2 (see also footnote 23). The unstated assumption of this statement is that a general agreement is universal and readily transparent to the experimental participants. I would not expect any agreement in this odd game to be plainly obvious across many individuals, the result of which is a high variance of observed outcomes.

²⁷ Oxoby and Spraggon (2008) and List (2007) also find that dictators are unwilling to take money that others have earned as a just desert.

Appendix B: Frequently Asked Questions/Proffered Comments and My Replies (Not necessarily for publication)

1. How are fair and fairness circularly used in economics?

We start with the observation of a fair outcome, a phenomenon that we wish to explain. To explain this fair outcome, we then mechanistically incorporate into the explanatory premises (the utility function) a social preference for that fair outcome. This is a circular process of explanation because the parameters added to the private preferences within an individual are unfortunately unobservable to those of us undertaking the task of explaining the phenomenon. We can't externally ground what we just added to the explanans. This is the reasoning process that I posed on page 1 to set the stage for the paper: "Why did proposer a offer $z = 5$ to b ?" "Because that's fair." "What is a fair outcome in the ultimatum game?" "Proposer a offers $z = 5$ to b ."

In other words, "we cannot ignore that within the logic of utility maximization the explanandum of *fair* is a perfunctory part of the explanans, i.e., a fair offer of $z = 5$ is part of a social preference for a fair outcome" (p. 3).

I will return to this below and show just how easy it is to slip into this circular line of reasoning.

2. The claims made about fairness terminology in English and other languages are misleading and, at times, wrong. Germans frequently also use the word *gerecht* in addition to *fair*, and French not only has a word for fairness, it has at least two: *juste* and *équitable*.

Let me be clear: I am not claiming that Anglophones are the only fair people on the planet. The claim in Wierzbicka (2006) is that "the word [*fair*] is one-to-one untranslatable into any other language" (p. 2). German speakers might note that they frequently also use the word *gerecht* in addition to fair; French speakers might say that they have not one but two words for it: *juste* and *équitable*. True enough. However, depending upon context, *gerecht* is also translatable into English as *just* or *equitable*. That's a three-to-one translation, not one-to-one, and probably one reason why the Germans directly import the English word. And while *équitable* is often translated as *fair*, it can also be rendered as *equitable*, and similarly, *juste* can be translated into English as *just* (and also *right*). These languages may have the sense of fairness, but we have a word for it that they do not. This presents us with an opportunity to pin down the concept of fairness as it relates to justice and equity.

3. The claim that *fair* is unique to English is not correct and seems to be based on the additional claim that the words *fair*, *just* and *equitable* mean different things in English such that *fair* occupies a particular place with special meanings.

Fletcher (1996), a well-known law professor, and Wierzbicka (2006), an eminent linguist, give serious scholarly consideration to the claim that "the word [*fair*] is one-to-one

untranslatable into any other language”. Given the breadth of languages that Wierzbicka has studied, a counterexample to this claim has not yet emerge. I am very clear in Section II to credit her for distinguishing the semantics of *fair* from *just* and *equitable* and for concluding that the concept is distinctly Anglo in origin. (By the way, Wierzbicka also reports that the original antonym of *fair* was *foul*, not *unfair*, as in a fair or foul ball in the rules of baseball or the British phrase “through means fair and foul”).

4. Although utility functions are private, why can't the payoff distributions that appear in their arguments be public, observable states of affairs?

While I do argue that utility functions are private, this does not preclude public, observable inputs to it. The thrust of my argument, however, is that we cannot separate the individual's social actions from the social motive, which is what specifying a utility function does. We can only discover an individual's motive from its tethered action. The utility function is a construction that equivocates when convenient on the distinction between motive and action. As I note in footnote 20: “What criteria can distinguish the actual actions taken by individual from the inner state of social preferences that consequently result in fairness? There are none?” See my replies to #5 and #6 below for a fuller description of this unconscious equivocation.

5. Can't social preference models just assume that people know what fairness is?

Explaining the phenomena of fairness with models that presume fairness is circular. The idea is to find and specify some other firmer ground, and Wierzbicka's definition provides the footing for us. More specifically, her “definition composed of semantic primes avoids circularity by appealing to two pieces of knowledge external to *b*—(1) what other people know about what others would think and do and (2) the limits thereof for actions that are bad” (p. 8). When faced with the inconvenient fact that we cannot observe the motive of fairness in the parameters of a utility function, we set that concern aside and assume for the moment that people know what it is so that the utility function only explains actions. However, when we want to then make predictions based upon a utility function, we conveniently forget that which we set aside and assume that the utility function is explicating the motive of fairness.

6. But integrating social concepts into the utility function has proven quite useful, including other concepts like efficiency and altruism. If we consider utility as a way of modeling behavior descriptively, relinquishing this option would represent a significant step backwards.

A related question from footnote 20 is “whether social preference models are meant for understanding the consequences of fairness or for explicating the cause of fairness. The problem is that this is a false dichotomy in describing what economists actually do when they ‘do science’. We do both.” This is part and parcel of the convenient equivocation I mentioned above. Sometimes we want to “explain many phenomena in one fell swoop”, but we also want to “make new predictions” (Camerer, p.101). Unfortunately, the

observed action and the unobservable motive are not separable and hence there's nothing external to ground this process of theoretical explaining or predicting.

7. If a concept is social and can only be applied within groups, why can't being social by definition be valued privately?

I am very precise in my paper. I do not assume that a person does not privately value being social. That is specifically one of the two false steps that I warn about in the introduction of Section III:

The goal of this section is to explicate the role that sociality plays in a fair experience vis-à-vis knowledge without taking either of two false steps: contending that specific social interactions are a necessary condition for a fair experience and dismissing sociality as not mattering at all.

I argue that *we* cannot in our rationalizing reconstruction of an individual's behavior use a private utility function to represent that private value because it severs social motives from social actions. That distinction is crucial.

8. Are you arguing that rules of fairness can be found?

Just as the mathematical rule that Wittgenstein's pupil follows is not publicly observable to his teacher, the rule of fairness that guides *a* to offer $z = 2$ to *b* is not publicly observable. What is publicly observable is the offer of $z = 2$, which is the *expression* of *a*'s rule of fairness. We literally only observe the offer, not the rule. I cannot overstate this distinction.

Apart from the obvious inability to peer inside the mind of another, why are the rules unformalizable? Because they are "not for use on a *single* occasion" (Wittgenstein 1980, §672, original emphasis); they must be flexible and spontaneous, and hence difficult to pin down, to contend with unforeseeable circumstances and an ever-changing background. Nevertheless, the activity is publicly observable, and every stage of it can be deemed by others to be fair or not. Sections III and IV confute the notion that the individual's expression of his social preferences as an action can be represented as a set of separable and private utilitarian "preferences" within him. In other words, to put first the idea of an inner social preference explaining or "doing" something that so-called "self-regarding" preferences do not, gets things the wrong way, for *it is only because an individual acts intelligibly within a pattern of social practice that we say some social motive takes place within him.*

Because the entire custom of social interactions is what matters, the problem boils down to one of *agreement* on the tacit social rules: Do the subjects agree on the contextual rules of the interaction? What is it about the social interaction that leads people to agree or to disagree on what is the right thing to do when faced *this* task?

9. Why should economists who construct social preference models be interested in a paper that targets their method?

Please note from the abstract that my goal in this paper is not to “target” the literature on social preferences. While I admit my attempt to attract readers with a Nietzschean-inspired title may provoke that mentality, the presentation of the paper is interrogatively therapeutic à la Wittgenstein. The deliberate purpose in the paper’s presentation is to lead the reader to look afresh at how we understand fairness in economics. That said, as I help the reader build this new picture, I am acutely aware that he or she views the concept through the firmly-rooted and preconceived notion of utility maximization, and hence, at times I will have to explicitly remind the reader the flaws of that familiar lens.

One goal of this paper is to begin a conversation on what we are doing when we model and experiment on fairness, because as this paper points out, it has deep consequences for how we interpret what we are learning from both valuable activities.

10. Why would an economics audience be interested in the approach of this paper?

Fair and *fairness* are frequently used words in everyday and professional discourse. According to the Econlit database at the time of my original submission, 635 journal articles in the last 15 years refer to *fairness* in their bibliographic record, with 334 of those just in the last 5 years. More to the point, 238 economics articles since 1992 use the word *fairness* in their title, 110 of which appeared in the last 5 years. Economic experiments are the chief engine and fuel for the profession’s interest in fairness.

This paper attempts to put the question marks deep down for understanding *fairness* in economic behavior. For example, how many economists realize that the word is distinctly Anglo in origin? As my paper points out, this must have implications for our understanding of a social concept that is becoming more and more relevant in policy and our daily lives. I think that it is difficult to argue that economists in general (and certainly experimental economists who are for the most part responsible for this interest in fairness) are not interested in what this paper has to say.

Again, while my goal is to move people away from thinking about fairness via social preference models, my goal is not purely a negative one. I am using Wittgenstein’s private language argument to construct a new perspective on fairness that is very different from a social preference model. I only directly address social preference models where I think the reader won’t let go of the utility maximization monocle.

To summarize, the case I am building in this paper is to turn around our whole examination of fairness. *The conclusion is not that we need a new explication nor that we need a new theory; what we need is a new frame of reference.*