Australasian Journal of Logic

Paraconsistent games and the limits of rational self-interest

Arief Daynes[*], Paraskevas Pagas[*], David Latimer[*], and Panagiotis Andrikopoulos[**]

[*] Department of Accounting and Finance, Portsmouth Business School, University of Portsmouth
[**] Department of Accounting and Finance, Leicester Business School, De Montfort University

Abstract

It is shown that logical contradictions are derivable from natural translations into first order logic of the description and background assumptions of the Soros Game, and of other games and social contexts that exhibit conflict and reflexivity. The logical structure of these contexts is analysed using proof-theoretic and model-theoretic techniques of first order paraconsistent logic. It is shown that all the contradictions that arise contain the knowledge operator K. Thus, the contradictions do not refer purely to material objects, and do not imply the existence of inconsistent, concrete, physical objects, or the inconsistency of direct sensory experience. However, the decision-making of rational self-interested agents is stymied by the appearance of such intensional contradictions. Replacing the rational self-interest axioms with axioms for an appropriate moral framework removes the inconsistencies. Rational moral choice in conflict-reflexive social contexts then becomes possible.

1 Introduction

In the Soros Game (a version of the Prisoner's Dilemma) there are three agents, the banker, and players a and b. Each player is given a card by the banker. On one side is written 'Give the other player \$5', and on the other 'Give me \$1'. The players do

not know each other or communicate with each other. At time $t = 0$ each player chooses one of the instructions to give to the banker, and the banker carries out the instructions so given one minute later at time $t = 1$. The game is played only once, the players are rational and self-interested, and the preceding description and background assumptions of the game are common knowledge. The *Soros Game Context* comprises the description of the game together with the background assumptions.

The simple, Soros Game version of the Prisoner's Dilemma is used here in order to emphasise the closed nature of the game. In many versions of the Prisoner's Dilemma there are unresolved questions concerning issues such as loyalty, revenge, the prior knowledge or assumptions of the players, etc., which can confuse the analysis. However, the Soros Game Context is closed, i.e. it is taken as part of the meaning of the game that there are no post-game consequences for the players beyond the financial gains the players receive from playing the game. Thus, there is no more to the game than is stated in the game context, and once the game is over it's over.

It is shown that natural translations into the language of first order logic of the Soros Game Context, and of other contexts such as the Surprise Examination Context and the Centipede Game Context, are inconsistent, where a sentence A (set of sentences S) is inconsistent if there is a formula B such that the logical contradiction $B \wedge \neg B$ is derivable from A (from some finite subset of sentences of S).

The base logics used in the formalisations of these game contexts are as follows. LP is extensional paraconsistent logic, the Logic of Paradox of Priest, 1979. Classical logic, CL, is obtained from LP by adding the rule Modus Ponens (MP) : $A, \neg A \vee B \Rightarrow B$. LPK is obtained from LP by adding $S5$-type modal rules for the knowledge operator K. CLK, equivalent to the classical system $S5$, is obtained from LPK by adding (MP). Thus, the logics are set up in such a way that the classical systems (paraconsistent systems) are obtained from the corresponding paraconsistent systems (classical systems) by adding (removing) (MP). Since a set of sentences S is inconsistent using the classical system CL (CLK) if and only if S is inconsistent in the corresponding paraconsistent system LP (LPK) (Theorems 2 and 3), both the classical and paraconsistent systems can be used in deriving the inconsistency theorems. However, it is technically much easier to use classical logic for this purpose, and in Theorem 4 it is proved that the Soros Game Context is inconsistent when the base logic is the classical system CLK. Once the Soros Game Context has been proved to be inconsistent, the base logic then switches to LPK so that the logical structure of the inconsistency can be investigated. Clearly the classical systems are of no use here, since every formula is derivable from an inconsistent set of sentences in classical logic. The main result here is that when LPK is used as base logic,

no inconsistencies in the realm of concrete physical objects are derivable (Theorem 5). For example, the inconsistency of the Soros Game Context does not imply that Player a both does and does not receive a payoff of \$5. The inconsistencies that arise always contain the knowledge operator K, and do not contradict the consistency of direct sensory experience.

Game theory has often been proposed as a logical foundation for economics, and as a framework allowing the unification of economics with the other social sciences (Gintis, 2000), i.e. game theory is often viewed as the particle physics and cosmology of economics. If the simple Soros Game Context is inconsistent, then it seems almost certain that virtually all other economic game contents which exhibit conflict and reflexivity are inconsistent too. Does this mean that game theory is bust? By no means. As will be seen from the analysis, the axioms that do the damage are those that formalise the neoclassical economics assumption of rational self-interest. However, while rational self-interest is assumed in classical game theory, it is explicitly excluded as an assumption in evolutionary game theory. These and other inconsistency theorems may well deliver a fatal blow to classical game theory. They are, however supportive of the evolutionary approach.

The term 'reflexivity' is used throughout the paper when discussing inconsistent games. It is not yet clear what is the root cause of the inconsistencies arising in game theory, and the use of terminology standard in mathematical logic, such as 'self-reference', 'vicious circle', 'impredicativity', etc. would prejudge the issue. For example, while there are clear analogies between the Soros Game Context and the Liar Paradox, the analogy is not exact; the Soros Game Context is not self-referential in any obvious sense. Thus, the suggestive, yet flexible term 'reflexivity' is used instead. 'Reflexivity' was coined in Soros (1987). The paper can be regarded as a formalisation within paraconsistent logic of some of the fundamental insights contained in Soros's work.

The plan of the paper is as follows. Section 2 clarifies some points on the authors' philosophical stance on paraconsistent logic and paraconsistent mathematics. Sections 3, 4 and 5 present the formal logical material. Section 6 proves the inconsistency of the Soros Game Context, and analyses the logical structure of this inconsistency. Section 7 interprets the derivations of the inconsistency of the Surprise Examination Context and of other backwards induction contexts of Priest (2000) within the paraconsistent interpretation presented here. The paper concludes in Section 8 with a review and discussion of the main results.

2 Paraconsistent mathematics

The derivation of a logical contradiction can be interpreted as a *reductio* proof that the purported entities do not in fact exist. For example, the proof of the inconsistency of the Russell Set, R, of all sets that are not members of themselves is sometimes taken as a proof that no such set exists. The alternative is to treat the proof of the inconsistency of R as a valid demonstration that inconsistent mathematical objects exist, an approach which has lead to the creation of the emerging field of paraconsistent mathematics. For paraconsistentists, inconsistent mathematical objects are perfectly valid generalisations of ordinary consistent mathematical objects, and denouncing such inconsistent objects as 'the work of the Devil' makes no more sense than similar denunciations of the negative integers or imaginary numbers in previous centuries.

The *reductio* interpretation is more compelling in paradoxes like that of the Barber of Seville, who shaves all those men in Seville who do not shave themselves, the paradox arising when it is asked who shaves the Barber. The obvious conclusion, given the implicit assumption that the Barber is a man, is that no such barber exists. However, in the case of games the *reductio* strategy seems not to work. For example, the Soros Game Context and other inconsistent game contexts apparently do exist, since there are people who appear to have rationally and self-interestedly played the Soros Game, while no one has ever been shaved by the Barber of Seville.

Inconsistent game contexts are of particular interest from the point of view of paraconsistent mathematics. The inconsistent objects that mathematicians (or at least those mathematicians actively working in the field of paraconsistent mathematics) have hitherto accepted as real, are the kinds of entities that are in some sense unobservable in principle, such things, for example as the Russell Set that occurs at the Absolute Infinite, the infinity that lies beyond all infinities. Inconsistent game contexts, however, bring paraconsistency down to Earth, since they arise in many of the ordinary, everyday activities of business and economic life which exhibit conflict and reflexivity.

However, paraconsistency cannot be brought down to Earth too far. Even the most confirmed defender of paraconsistent mathematics baulks at the idea of, for example an inconsistent beach ball that is red all over and green all over at the same time. Such a concrete example of an inconsistent entity, contradicting the consistency of direct sensory experience, could be accepted only at the expense of radically and perversely changing the ordinary meanings of words. It appears to be accepted by everyone that inconsistencies do not arise in the realm of the ordinary concrete physical objects of everyday experience. However, game contexts, and other

social constructs such as human-made laws, regulations, conventions and linguistic entities, are abstract entities, not concrete objects, and the existence of inconsistent game contexts does not imply the existence of inconsistent concrete objects. For example, if two economics students Allie and Bobbie begin to play the Soros Game, say, at 10.00 a.m., sitting at a certain table in a certain room in a certain college building in a certain town, it does not follow from the inconsistency of the Soros Game Context that Bobbie both has and does not have \$5 in her hand at the completion of the game. Thus, 'At 10.01 a.m. the banker is handing Bobbie a \$5 banknote' is a sentence that is unambiguously true or unambiguously false, not both true and false. The money held by Bobbie at the end of the game is, like the beach ball, a concrete entity where no inconsistency can arise. What is inconsistent is the sentence (an abstract object) describing the Soros Game Context, the sentence itself modelling such abstract concepts as 'knowledge', 'rationality', 'self-interest' and 'dominant strategy'.

That the inconsistency of abstract game contexts does not imply the inconsistency of concrete physical entities follows from the results in the following sections, where the structure of the inconsistency of the Soros Game Context is analysed using results from the proof theory and model theory of paraconsistent logic.

3 Extensional paraconsistent logic

Sections 3, 4 and 5 present enough of the formal development of paraconsistent and classical logic to give rigorous proofs of the technical results presented in the paper. The notation, definitions and theorems given in Sections 3, 4 and 5 are fairly standard. Readers can skim these sections to fix the notation and move on to Section 6.

An extensional first order language, L, comprises countably infinitely many individual variables, $x, y, z, \ldots$, the identity relation $=$, connectives $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), the quantifiers $\forall$ (universal quantifier) and $\exists$ (existential quantifier), and brackets (,). These comprise the logical symbols of L, and are common to all first order languages. L may contain, in addition certain non-logical symbols, which depend on the application. The non-logical symbols comprise individual constant symbols, a, b, c, $\ldots$, n-place relation symbols, $n \geqslant 1$, P^n, Q^n, R^n, ..., and n-argument function symbols, $n \geqslant 1$, f^n, g^n, h^n, $\ldots$. The particular standard conventions used in this paper for building terms and formulas of L from the symbols of L will be evident from the following development. It is noted, however that $A \rightarrow B$ (material implication) is an abbreviation for $(\neg A) \vee B$ and that $A \leftrightarrow B$ (material equivalence) is an abbreviation for $(A \rightarrow B) \wedge (B \rightarrow A)$.

The *closed Hilbert formulation of extensional paraconsistent logic*, the *Logic of Paradox*, LP, has the following axioms, rules and meta-rules.

Logical Axioms:

(A1) Every sentence of the form $(\forall x_0)...(\forall x_n)(A \vee \neg A)$ is a logical axiom of LP.

(A2) Every sentence of the form $(\forall x_0)...(\forall x_n)t = t$ is a logical axiom of LP.

Logical Inference Rules:

(R1) $A, B \Rightarrow A \wedge B$ (from A and B to infer $A \wedge B$)

(R2) $A \wedge B \Rightarrow A$ (from $A \wedge B$ to infer A)

(R3) $A \wedge B \Rightarrow B$

(R4) $A \Rightarrow A \vee B$

(R5) $B \Rightarrow A \vee B$

(R6) $A \Leftrightarrow \neg\neg A$ (Double Negation Law)

(R7) $\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$

(R8) $\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$ ($(R7)$ and $(R8)$ are the De Morgan Laws)

(R9) $A \wedge B \Leftrightarrow B \wedge A$

(R10) $A \vee B \Leftrightarrow B \vee A$ ($(R9)$ and $(R10)$ are the Commutative Laws)

(R11) $A \wedge (B \wedge C) \Leftrightarrow (A \wedge B) \wedge C$

(R12) $A \vee (B \vee C) \Leftrightarrow (A \vee B) \vee C$ ($(R11)$ and $(R12)$ are the Associative Laws)

(R13) $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$

(R14) $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$ ($(R13)$ and $(R14)$ are the Distributive Laws)

(R15) $A \Rightarrow A \wedge A$

(R16) $A \vee A \Rightarrow A$

(QR1) $(\forall x)A \Rightarrow A(t/x)$

(QR2) $A(t/x) \Rightarrow (\exists x)A$

(QR3) $(\forall x)\neg A \Leftrightarrow \neg(\exists x)A$

(QR4) $(\exists x)\neg A \Leftrightarrow \neg(\forall x)A$

(QR5) $A \Rightarrow (\forall x)A$

(QR6) $(\exists x)A \Rightarrow A$

(QR7) $(\forall x)(A \vee B) \Rightarrow (\forall x)A \vee B$

(QR8) $(\exists x)A \wedge B \Rightarrow (\exists x)(A \wedge B)$

(IR1) $s = t \wedge A \Rightarrow B$, where B is obtained from A by replacing some free occurrences of s with free occurrences of t.

In $(QR1)$ and $(QR2)$ the term t is free for x in A. In $(QR5)$ and $(QR6)$ x does not occur free in A. In $(QR7)$ and $(QR8)$ x does not occur free in B.

Meta-rules:

(MR1) If $A \Rightarrow B$ is a rule then $A \wedge C \Rightarrow B \wedge C$ is a rule.

(MR2) If $A \Rightarrow B$ is a rule then $A \vee C \Rightarrow B \vee C$ is a rule.

(MR3) If $A \Rightarrow B$ is a rule then $(\forall x)A \Rightarrow (\forall x)B$ is a rule.

(MR4) If $A \Rightarrow B$ is a rule then $(\exists x)A \Rightarrow (\exists x)B$ is a rule.

The *closed Hilbert version of classical logic*, CL, is obtained from LP by adding just one rule scheme, *Modus Ponens* $(MP) : A, \neg A \vee B \Rightarrow B$, for arbitrary formulas A and B.

Let S be a set of formulas and A a formula. A *derivation of A from S* in LP is a sequence of formulas $A_1, ..., A_n = A$ such that for each $i = 1, ..., n$:

1. A_i is a logical axiom of LP or an element of S, or
2. A_i follows from A_j and A_k by $(R1)$, for some $j, k < i$, or
3. A_i follows from A_j by one of the other logical rules of LP.

If A is deducible from S then A is said to be a *syntactic consequence* of S, written $S \vdash A$. For CL the definition is the same except that in (2) A_i may also follow from A_j and A_k by (MP), for some $j, k < i$.

4 Soundness and adequacy meta-theorems

Note that the following semantics is given in terms of the diagram of the model, i.e. the set of atomic and negated atomic formulas that hold in the model, rather than in terms of a truth function from atomic formulas into the truth values $\{0\}$ (false), $\{1\}$ (true), and $\{0, 1\}$ (true and false), as in Priest's semantics for LP (Priest, 1979, 2006(a)). While both approaches give essentially the same technical results, the diagram approach used here gives a somewhat smoother development overall for the present applications and intended sequels.

An *LP model for a language* L is an ordered pair $M =< D, I >$, where D is a non-empty domain, and I, the *interpretation function*, is a function from the set of all individual constant symbols, relation symbols and function symbols of L such that:

1. $I(c)$ is an element of D for each individual constant symbol c.

2. For each n-place relation symbol P^n other than the identity relation symbol, $=$, $I(P^n)$ is a set of expressions of the form $P^n(d_1, \ldots, d_n)$ or $\neg P^n(d_1, \ldots, d_n)$, where $d_1, \ldots., d_n$ are elements of D, such that at least one of $P^n(d_1, \ldots, d_n)$, $\neg P^n(d_1, \ldots, d_n)$ is in $I(P^n)$ for every $d_1, \ldots, d_n$ in D.

3. $I(=)$ is a set of expressions of the form $d_1 = d_2$ or $\neg d_1 = d_2$, where d_1 and d_2 are elements of D, such that:

   (a) $d_1 = d_2$ is in $I(=)$ if and only if d_1 is identical to d_2, and

   (b) If d_1 is not identical to d_2 then $\neg d_1 = d_2$ is in $I(=)$.

4. $I(f^n)$ is an n-argument function from D^n to D for each n-argument function symbol f^n.

A valuation v on M is a function from the set of individual variables of L such that $v(x)$ is an element of D for each variable x. The valuation v is extended recursively to a function on all terms by $v(f^n(t_1, \ldots, t_n)) = I(f^n)(v(t_1), \ldots, v(t_n))$, for terms $f^n(t_1, \ldots, t_n)$. If A is a formula then the satisfaction of A at v in M, (i.e. the truth of A in M with respect to the valuation v), $v \vDash_M A$, is given by the following recursive definition.

1. If $A(t_1, \ldots, t_n)$ is an atomic or negated atomic formula then $v \vDash A$ iff $A(v(t_1), \ldots, v(t_n))$ is in $I(A)$.

2. $v \vDash A \wedge B$ iff $v \vDash A$ and $v \vDash B$.

3. $v \vDash A \vee B$ iff $v \vDash A$ or $v \vDash A$.

4. $v \vDash \neg(A \wedge B)$ iff $v \vDash \neg A$ or $v \vDash \neg B$.

5. $v \vDash \neg(A \vee B)$ iff $v \vDash \neg A$ and $v \vDash \neg B$.

6. $v \vDash \neg\neg A$ iff $v \vDash A$.

7. $v \vDash (\forall x)A$ iff $v' \vDash A$ for every valuation v' that differs from v only in that v' may take a different value of D at x. Such a v' is called an x-variant of v.

8. $v \vDash (\exists x)A$ iff $v' \vDash A$ for some x-variant v' of v.

9. $v \vDash \neg(\forall x)A$ iff $v' \vDash \neg A$ for some x-variant v' of v.

10. $v \vDash \neg(\exists x)A$ iff $v' \vDash \neg A$ for every x-variant v' of v.

A is a semantic consequence of S, $S \vDash A$, if for every model M and every valuation v in M, $v \vDash_M B$ for every formula B in S implies $v \vDash_M A$. A sentence A holds, or is true in a model M if $v \vDash_M A$ for every valuation v in M; otherwise A fails, or is false in M.

A classical model (CL model, or consistent model) is defined as for an LP model except that clauses (1) and (2)(b) of the definition of the interpretation function I are changed to (1)′ and (2)(b)′, as follows.

(1)′ For each n-place relation symbol P^n other than the identity relation symbol, $=$, $I(P^n)$ is a set of expressions of the form $P^n(d_1, \ldots, d_n)$ or $\neg P^n(d_1, \ldots, d_n)$, where $d_1, \ldots, d_n$ are elements of D, such that *exactly one* of $P^n(d_1, \ldots, d_n)$, $\neg P^n(d_1, \ldots, d_n)$ is in $I(P^n)$ for every $d_1, \ldots, d_n$ in D.

(2)(b)′ $I(=)$ is a set of expressions of the form $d_1 = d_2$ or $\neg d_1 = d_2$, where d_1 and d_2 are elements of D, such that d_1 is not identical to d_2 *if and only if* $\neg d_1 = d_2$ is in $I(=)$.

Theorem 1. *(Soundness and Adequacy Meta-Theorems for LP and CL) Let A be a formula and let S be a set of formulas in the language L. Then $S \vdash_{LP} A$ ($S \vdash_{CL} A$) if and only if $S \vDash_{LP} A$ ($S \vDash_{CL} A$).*

Proof. Soundness and adequacy meta-theorems for the closed Hilbert formulations of paraconsistent and classical logic of this paper are given in the Appendix of Daynes (2000) (where LP is called CPQ and CL is called CLQ). □

Although LP is much weaker than CL, it is nevertheless the case that if S is inconsistent in CL then it is also inconsistent in LP.

Theorem 2. *(a) Let A be derivable from S in CL. Let $B_0, B_0 \rightarrow C_0, \ldots, B_n, B_n \rightarrow C_n$ be all the antecedents of applications of (MP) used in this derivation. Then $A \vee (B_0 \wedge \neg B_0) \vee \ldots \vee (B_n \wedge \neg B_n)$ is derivable from S in LP. (b) If S is inconsistent in CL then it is inconsistent in LP.*

Proof. (Outline) (a) If $B_i, B_i \rightarrow C_i \Rightarrow C_i$ is a step in the CL derivation of A from S, replace the step with the derived LP rule: $B_i, B_i \rightarrow C_i \Rightarrow C_i \vee (B_i \wedge \neg B_i)$. The extra disjunct $(B_i \wedge \neg B_i)$ is then carried along as a passenger in the rest of the derivation. (b) By (a), if a logical contradiction $A \wedge \neg A$ is derivable from S in CL then a formula, C, of the form $(A \wedge \neg A) \vee (B_0 \wedge \neg B_0) \vee \ldots \vee (B_n \wedge \neg B_n)$ is derivable from S in LP. $\neg C$ is a logically valid formula of CL, and it is a well-known fact that all classically valid formulas are derivable in LP. Therefore, the logical contradiction $C \wedge \neg C$ is derivable from S in LP. □

5 Extensional and intensional logics

To formally model the rational self-interest assumption it is necessary to introduce intensional knowledge operators, K, where for a sentence A, KA is interpreted as 'It is known that A'. In general there will be more than one K operator. For example, the intended interpretation of $K_{(a,t)}A$ may be that A is known by a at time t. However, in the analysis of the Soros Game only a single K operator is required. This is because the context of the game is common knowledge, and because the Soros Game is a 1-step game, so that there is no need to distinguish what is known at different times. Thus KA is interpreted as 'It is known (by everyone) at the start of the game (at time $t = 0$) that A'.

The logic CLK is obtained from CL by extending the definitions of Section 3 as follows. The language of CL is expanded by adding a single knowledge operator K, the formation rules for constructing formulas are extended by adding the formation rule: If A is a formula then KA is a formula, and the following new logical axioms and rules are adjoined.

(A3) If A is a logical axiom then KA is a logical axiom.

(KR1(a)) $KA \Rightarrow A$

(KR1(b)) $\neg A \Rightarrow \neg KA$

(KR2(a)) $KA, KB \Rightarrow K(A \wedge B)$

(KR2(b)) $\neg K(A \wedge B) \Rightarrow \neg KA \vee \neg KB$

(KR3(a)) $KA \Rightarrow KKA$

(KR3(b)) $\neg KKA \Rightarrow \neg KA$

(KR4(a)) $\neg K\neg A \Rightarrow K\neg K\neg A$

(KR4(b)) $\neg K\neg KA \Rightarrow KA$

(MRK) If $A \Rightarrow B$ is a logical rule then $KA \Rightarrow KB$ is a logical rule.

(KMP) $KA, K(A \rightarrow B) \Rightarrow KB$

($A3$) says that rational agents know the axioms of logic, and ($KR2(a)$), (KMP) and (MRK) say that rational agents know the logical consequences of known propositions. ($KR1(a)$) says that known propositions are true. ($KR3(a)$) formalises the common knowledge assumption, since it says 'If A is known (by everyone) then it is known (by everyone) that A is known (by everyone)'. ($KR4(a)$) says that if rational agents know that A might be true, i.e. that A is not known to be false, then this fact is known. ($KRi(b)$) is the contrapositive of ($KRi(a)$), for $i = 1, 2, 3$ and 4.

These new axioms and rules formalise a stronger notion of knowledge and rationality than is actually required in the inconsistency proofs. In particular, the rules ($KR4(a)$) and ($KR4(b)$) are not used anywhere in the inconsistency proofs, as is evident from the proofs themselves. However, it is convenient to include ($KR4(a)$) and ($KR4(b)$) since it gives a logic with a much simpler formal semantics.

Paraconsistent modal logic, LPK is obtained from CLK by deleting (MP) and rule (KMP), the K version of Modus Ponens.

Note that CLK is just a reformulation of the first order modal logic $S5$ with identity. Such a reformulation is necessary. It would be awkward to use a standard formulation of $S5$, since deleting (MP) from such a formulation would not give LPK.

The semantics for CLK is the usual constant domain possible worlds semantics, where the accessibility relation R is an equivalence relation. Thus, a CLK model, M, of the first order language L can be taken to be a non-empty indexed set of CL models $M = \{M_i =< D, I_i >: i \in J\}$, where each M_i is a CL world as in Section 4, such that the domain D of each world M_i is the same for each $i \in J$, and such that each interpretation function I_i, $i \in J$, assigns the same element of D to each individual constant symbol of L, and the same function to each function symbol of L. For the identity relation symbol $=$, $I_i(=) = \{d = d : d \in D\} \cup \{\neg d_1 = d_2 : d_1, d_2 \in D$

and d_1 not identical to $d_2\}$, for each $i \in J$. The relations assigned to the non-logical relation symbols of L, however depend on the index i. A valuation, v, again assigns an element $v(x)$ of D to each individual variable x. The satisfaction of a formula A at v at world M_i, $v \vDash_{M(i)} A$ is as for Section 4, extended with the clause: For every i in J, $v \vDash_{M(i)} KA$ iff $v \vDash_{M(j)} A$ for every $j \in J$. A sentence A is true in the model M if every valuation v satisfies A in the distinguished base world M_0.

The definitions for LPK are as for CLK except that now the M_i are taken to be LP worlds, and the following clause for negated K formulas is added: For every i in J, $v \vDash_{M(i)} \neg KA$ iff for some $j \in I$, $v \vDash_{M(j)} \neg A$.

While adequacy results are well-known for the classical first order modal logics, the situation for paraconsistent first order modal logics is less clear. Priest (2008, Appendix 11(a)) proves adequacy for the propositional fragments of various paraconsistent modal logics. While the extension of these results to the first order case appears to be straightforward, there is currently no published proof in the literature. However, for the purposes of this paper only the soundness of LPK with respect to the semantics, and the following analogue of Theorem 2 are required.

Theorem 3. *(a) Let $A_0, \ldots, A_n$ be a derivation of A_n from S in CLK. Let $K^{n(0)}B_0$, $K^{n(0)}(B_0 \rightarrow C_0), \ldots, K^{n(k)}B_k, K^{n(k)}(B_k \rightarrow C_k)$ be all the antecedents of applications of (MP) used in this derivation (including the K versions of (MP) given by the rule (KMP), where $K^0A =^{df} A$, and $K^{n+1}A =^{df} KK^nA$). Then $A_n \vee (B_0 \wedge \neg B_0) \vee \ldots \vee (B_k \wedge \neg B_k)$ is derivable from S in LPK. (b) If S is inconsistent in CLK then it is inconsistent in LPK.*

Proof. (Outline) (a) The basic trick behind the proof is the same as in Theorem 2. The details of the formal proof by induction on the length n of the derivation of A_n from S are omitted. (b) follows directly from (a). □

6 Paraconsistent analysis of the Soros Game

The Soros Game Context is defined as in Section 1. The standard non-formal analysis of the Soros Game Context is as follows. Since the players a and b each have the choices 'Give the other player \$5' and 'Give me \$1', considering a's position it is clear that a will be better off choosing to take \$1 whatever choice is made by b. Similarly it is clear that b will be better off choosing to take \$1, whatever choice is made by a. Thus, (\$1, \$1) is a Nash equilibrium of the game, and moreover, choosing to take \$1 is a dominant strategy for each player. Therefore, the optimal strategies are for each player to choose 'Give me \$1' at time $t = 0$, when they will receive \$1 each at time

$t = 1$. The puzzling aspect of the game is that both players would have been better off if each had chosen to give \$5, when they would each have received \$5. Despite this the standard position in game theory is that the rational strategy for each player is to choose to take \$1.

It will now be shown that the following formalisation of the Soros Game in first order logic is inconsistent.

The language of the Soros Game Context comprises:

The epistemic operator K (There is no need to have different K operators for the banker and for the different players, since all the axioms of the Soros Game Context are common knowledge. Also, since the Soros Game is a 1-step game there is no need to distinguish knowledge at different times. Thus, only one K operator is required, with KA meaning 'It is known by everyone, at the beginning of the game at time $t = 0$, that A').

Individual constant symbols T, G (the choices made at time $t = 0$, where T is to take \$1 from the banker, and G is to instruct the banker to give \$5 to the other player), 0, 1, 5 and 6 (the numerals for the numbers zero, one, five and six). In the interests of parsimony there is no need to introduce individual constant symbols for the players a and b or the banker.

1-place relation symbols PC (where '$PC(x)$' means that x is a permissible choice at time $t = 0$), C_a, C_b (where '$C_i(x)$' means that player i chooses x at time $t = 0$, $i = a, b$), P_a, P_b (where '$P_i(x)$' means that player i receives payoff x at time $t = 1$, $i = a, b$).

2-place relation symbol $<$ (where $<$ denotes the ordinary linear ordering of the natural numbers, with $0 < 1 < 5 < 6$).

The axioms of the Soros Game Context are given next, each axiom being followed by a non-formal English translation or explanation:

Axiom (1) If c and d are individual constant symbols from $\{T, G, 0, 1, 5, 6\}$ then:

1. If c and d are distinct such symbols then $\neg c = d$ is an axiom.

2. If c and d are from the set of numerals $\{0, 1, 5, 6\}$ and the numbers $\acute{c}$ and $\acute{d}$ corresponding to c and d are such that $\acute{c}$ is less than $\acute{d}$ then $c < d$ is an axiom; otherwise $\neg c < d$ is an axiom.

3. $(\forall x)(x = T \vee x = G \vee x = 0 \vee x = 1 \vee x = 5 \vee x = 6)$.

By Axiom (1) the domain of discourse contains exactly the six distinct objects T, G, 0, 1, 5 and 6, and the usual $<$ relation holds between the numbers 0, 1, 5 and 6.

Axiom (2) $(\forall x)(PC(x) \leftrightarrow x = T \vee x = G)$.

The permissible choices for the players at time $t = 0$ are exactly T and G.

Axiom (3) $(C_a(T) \vee C_a(G)) \wedge (C_b(T) \vee C_b(G))$.

Players a and b each must make at least one of the permissible choices, T (take \$1) or G (give \$5).

Axiom (4) $\neg(C_a(T) \wedge C_a(G)) \wedge \neg(C_b(T) \wedge C_b(G))$.

No player can choose both T and G, i.e. at most one choice is permitted.

By Axioms (2), (3) and (4) each player makes exactly one choice from $\{T, G\}$.

Axiom (5) $(C_a(T) \wedge C_b(T) \rightarrow P_a(1) \wedge P_b(1)) \wedge (C_a(T) \wedge C_b(G) \rightarrow P_a(6) \wedge P_b(0)) \wedge (C_a(G) \wedge C_b(T) \rightarrow P_a(0) \wedge P_b(6)) \wedge (C_a(G) \wedge C_b(G) \rightarrow P_a(5) \wedge P_b(5))$.

This gives the payoffs to players a and b at time $t = 1$ depending on the choices made by a and b at time $t = 0$.

Axiom (6) $(\forall x)(\forall y)(P_a(x) \wedge P_a(y) \rightarrow x = y) \wedge (\forall x)(\forall y)(P_b(x) \wedge P_b(y) \rightarrow x = y)$.

No player can receive two different payoffs.

By the above axioms each player makes exactly one choice at time $t = 0$, and receives exactly one payoff at time $t = 1$, these payoffs being determined by the choices made by the players at the start of the game at time $t = 0$ in accordance with the payoff rules of the Soros Game.

Axiom (7) $(\forall x)(\forall y)(\forall z)(\forall w)(K(PC(x) \wedge PC(z) \wedge (C_a(x) \rightarrow P_a(y)) \wedge (C_a(z) \rightarrow P_a(w)) \wedge w < y) \rightarrow C_a(x)) \wedge (\forall x)(\forall y)(\forall z)(\forall w)(K(PC(x) \wedge PC(z) \wedge (C_b(x) \rightarrow P_b(y)) \wedge (C_b(z) \rightarrow P_b(w)) \wedge w < y) \rightarrow C_b(x))$.

If it is known that the payoff y to a at time $t = 1$ of permissible choice x at time $t = 0$ is greater than the payoff w to a at time $t = 1$ of permissible choice z at time $t = 0$, then a chooses x at time $t = 0$, and similarly for b. The use of material implication in this axiom requires further justification, as given below.

Axiom (8) $(\forall x)(PC(x) \wedge K(\forall y)(\forall z)(\forall w)(\forall u)(PC(y) \wedge PC(w) \wedge ((C_a(x) \wedge C_b(y) \rightarrow P_a(z)) \wedge (C_a(w) \wedge C_b(y) \rightarrow P_a(u)) \wedge \neg x = w \rightarrow u < z)) \rightarrow C_a(x))) \wedge$

$(\forall x)(PC(x) \wedge K(\forall y)(\forall z)(\forall w)(\forall u)(PC(y) \wedge PC(w) \wedge ((C_b(x) \wedge C_a(y) \rightarrow P_b(z)) \wedge (C_b(w) \wedge C_a(y) \rightarrow P_b(u)) \wedge \neg x = w \rightarrow u < z) \rightarrow C_b(x)))$.

If it is known that the payoff, z, to a of permissible choice x is higher than the payoff, u, of permissible choice w, for every permissible choice w different from x, for any permissible choice, y, by b, then a chooses x, i.e. if it is known that choosing x is a dominant strategy for a then a chooses x, and similarly for b. This axiom too is given further justification below.

Axiom (9) If *Axiom* (n), $n = 1, ..., 8$, is an axiom of the Soros Game Context then K(*Axiom* (n)) is an axiom of the Soros Game Context.

This, together with the K rules of CLK formalises the common knowledge assumption.

The formalisation of the Soros Game Context, (SGC), is defined to be the conjunction of *Axiom* (1) to *Axiom* (9).

Axioms (1) to (6) are self-evidently valid formalisations of very basic aspects of the Soros Game Context and appear to be indisputable. *Axioms* (7) and (8), which formalise aspects of the rational self-interest assumption, appear to be unobjectionable provided $\rightarrow$ is taken to be the indicative implication of ordinary English (or maybe some other ordinary language implication, such as the subjunctive conditional). However, $\rightarrow$ stands for material implication, and is it generally accepted that material implication is not an adequate formalisation of the ordinary language indicative implication (and almost universally accepted that it is not an adequate formalisation of the ordinary language subjunctive conditional). Therefore, *Axioms* (7) and (8) require further justification beyond the merely suggestive ordinary language translations that follow the statements of the axioms.

Firstly, it is accepted that *Axioms* (7) and (8) do *not* fully capture the meaning of the informal translations immediately following the statements of these axioms. Indeed, it is clear that the most complete and accurate formalisations of economic game contexts require the use of intensional implications, i.e. implications which, unlike the extensional material implication, require some kind of many-worlds semantics for their formal analysis. While there is considerable debate as to the correct identification, classification, analysis and formalisation of the appropriate intensional implications, it is certainly the case that game-theoretic contexts are intensional in nature.

However, it is not required that the material implication forms of *Axioms* (7) and (8) capture the full meaning of their ordinary language translations. While it is accepted that (SGC) fails to capture the full meaning of the non-formal Soros Game

Context, for the results of this paper it is sufficient only that the axioms of (SGC) *be true*, given the intended interpretation of the formal symbols.

Perhaps the best way to see the truth of *Axiom* (7) is to take its negation, $\neg(Axiom\ (7))$, and to observe that a certain consequence of $\neg(Axiom\ (7))$, the sentence $(*)$ below, is clearly false of rational self-interested agents in the Soros Game Context.

$\neg(Axiom\ (7))$ is the disjunction: $(\neg(Axiom\ (7)))(a) \vee (\neg(Axiom\ (7)))(b) =^{df}$

$(\exists x)(\exists y)(\exists z)(\exists w)(K(PC(x) \wedge PC(z) \wedge (C_a(x) \rightarrow P_a(y)) \wedge (C_a(z) \rightarrow P_a(w)) \wedge w < y) \wedge \neg C_a(x)) \vee (\exists x)(\exists y)(\exists z)(\exists w)(K(PC(x) \wedge PC(z) \wedge (C_b(x) \rightarrow P_b(y)) \wedge (C_b(z) \rightarrow P_b(w)) \wedge w < y) \wedge \neg C_b(x))$.

From the first disjunct, $(\neg(Axiom\ (7)))(a)$, the following sentence, $(*)$, is derivable in CLK, making use of *Axioms* $(1)-(6)$ together with the K-versions of *Axioms* $(1)-(6)$:

$$(*)\ (\exists x_1)(\exists y_1)(\exists x_2)(\exists y_2)(K((C_a(x_1) \wedge P_a(y_1)) \vee (C_a(x_2) \wedge P_a(y_2))) \wedge K(y_1 < y_2) \wedge K(PC(x_1)) \wedge K(PC(x_2)) \wedge C_a(x_1))$$

Non-formally, in $(*)$ Agent a, i.e. Allie, says to herself: 'I know that I will choose x_1 at $t = 0$ and receive $\$y_1$ at $t = 1$, or I will choose x_2 at $t = 0$ and receive $\$y_2$ at $t = 1$. In addition I know that $y_1 < y_2$ and that both x_1 and x_2 are choosable, i.e. that both x_1 and x_2 are permissible choices, and yet I will choose x_1, i.e. I know that choosing x_1 gives a smaller payoff than choosing x_2, and yet I will choose x_1'. But this is not how Allie would actually reason if she were acting rationally and self-interestedly in the context of the Soros Game.

In the above argument the closed nature of the game becomes important. $(*)$ is false for a rational self-interested Allie, because choosing x_2 and receiving the higher amount of $\$y_2$ has no consequences beyond the purely financial outcomes as given by the game. For example, Allie's choice does not provide knowledge useful to the other player in future social interactions that would have adverse consequences for Allie herself. Once the game is over it's over.

Since $(*)$ follows from $(\neg(Axiom\ (7)))(a)$, and $(*)$ is false of rational self-interested players in the closed context of the Soros Game, $(\neg(Axiom\ (7)))(a)$ is false. By an exactly analogous argument the second disjunct of $\neg(Axiom\ (7))$, $(\neg(Axiom\ (7)))(b)$, is also false. Thus, $\neg(Axiom\ (7))$ is false of rational self-interested agents in the closed Soros Game Context, and so *Axiom* (7) must be the one that is true.

A similar analysis of the negation of *Axiom* (8), $\neg(Axiom\ (8))$, shows that *Axiom* (8) is true under the intended interpretation of the formal symbols, since $\neg(Axiom\ (8))$ says, in the presence of *Axioms* $(1)-(6)$ together with the K-versions of *Axioms*

$(1)-(6)$, that at least one of the players does not choose a known dominant strategy, an assertion that is clearly false.

Finally, *Axiom* (9) expresses the non-controversial fact that the Soros Game Context is known by both players.

Theorem 4. *The Soros Game Context is inconsistent.*

Proof. By Theorem 3(b), to show that the Soros Game Context is inconsistent, the classical system CLK can be used as base logic. It is shown that there is a logical contradiction $\perp$ that is a CLK-semantic consequence of the Soros Game Context. By the adequacy meta-theorem for CLK (i.e. for the classical modal logic $S5$), $\perp$ is CLK-syntactic consequence of the Soros Game Context. Thus, the Soros Game Context is inconsistent. Let $M = \{M_i : M_i =< D, I_i >, i \in J\}$ be a CLK model with base world M_0 such that all the axioms of the Soros Game Context are true in M, i.e. such that all the axioms are true at the base world M_0. By *Axioms* (1) to (6) it is clear that choosing T is a dominant strategy for each player. By *Axiom* (9), *Axioms* (1) to (6) are known, and by the K rules of CLK the logical consequences of known axioms are also known. Thus, it is known that choosing T is a dominant strategy for each player. By *Axiom* (8) both a and b choose T. Thus $C_a(T) \wedge C_b(T)$ holds at M_0. Again, by the K rules of CLK, $K(C_a(T) \wedge C_b(T))$ holds at M_0. Since, by *Axiom* (4), each player can make only one choice $(*) : K(\forall x)(C_a(x) \leftrightarrow C_b(x))$ holds at M_0. Therefore, $(**) : K(C_a(T) \leftrightarrow C_b(T)) \wedge K(C_a(G) \leftrightarrow C_b(G))$ holds at M_0. By $(**)$ and *Axiom* (7), $C_a(G)$, i.e. since it is known that a and b make the same choice, a choosing G determines that b also chooses G, which gives a higher payoff to a than a (and hence b also) choosing T. Therefore a chooses both T and G, contradicting *Axiom* (4). Thus, take $\perp$ to be $C_a(T) \wedge \neg C_a(T)$. □

The next theorem shows that no Beach Ball type inconsistencies are derivable from the Soros Game Context.

Theorem 5. *For no K-free formula A is it the case that $A \wedge \neg A$ is derivable from the Soros Game Context (SGC) using LPK as base logic, i.e. no concrete inconsistencies are derivable from the Soros Game Context in paraconsistent logic.*

Proof. Consider the following LPK model, M, of (SGC). $M = \{M_0 =< D, I_0 >, M_1 =< D, I_1 >\}$, with base world M_0. D = the set of symbols $\{T, G, 0, 1, 5, 6\}$. For each individual constant symbol c of the language of the Soros Game Context, $I_i(c) = c$, $i = 0, 1$. M_0 is taken to be a consistent LP world in which *Axioms* (1) to (6) are true, and in which $PC(T)$, $PC(G)$, $C_a(T)$, $C_b(T)$, $P_a(1)$ and $P_b(1)$ hold, i.e. M_0 is a consistent LP world in which a and b each choose the dominant strategy, and

each receive a payoff of \$1. These specifications determine the values of I_0 exactly. M_1 is exactly the same as M_0, except that $C_a(G)$, $C_b(G)$, $\neg C_a(T)$ and $\neg C_b(T)$ also hold at M_1, so that M_1 is over determined, or inconsistent. To see that M is a model of (SGC) it is first be shown that *Axioms* (1) to (8) hold at both world M_0 and world M_1. By the clause for the satisfaction of a formula of the form KB at a world, it then follows that $K(Axiom\ (n))$ holds at each world, for each $n = 1, ..., 8$. Thus, all the axioms of (SGC) hold at the base world M_0, and M is a model of (SGC). Clearly the K free axioms, *Axioms* (1) to (6) hold at both M_0 and M_1. To see that *Axiom* (7) holds at each world, consider the first conjunct of *Axiom* (7), *Axiom* (7)(a): $(\forall x)(\forall y)(\forall z)(\forall w)(K(PC(x) \wedge PC(z) \wedge (C_a(x) \rightarrow P_a(y)) \wedge (C_a(z) \rightarrow P_a(w)) \wedge w < y) \rightarrow C_a(x))$. This is of the form: $(\forall x)(\forall y)(\forall z)(\forall w)(KA \rightarrow C_a(x))$, which is LPK equivalent to $(*)$: $(\forall x)(\forall y)(\forall z)(\forall w)(\neg KA \vee C_a(x))$. A valuation v satisfies $\neg KA$ at a world M_i if and only if v satisfies $\neg A$ at some world M_j. It is shown that $\neg A$ holds at M_1 for all choices of x, y, z and w, and hence that $(*)$ holds at both M_0 and M_1. $\neg A$ is $\neg(PC(x) \wedge PC(z) \wedge (C_a(x) \rightarrow P_a(y)) \wedge (C_a(z) \rightarrow P_a(w)) \wedge w < y)$, which is LPK equivalent to $(**)$: $\neg PC(x) \vee \neg PC(z) \vee (C_a(x) \wedge \neg P_a(y)) \vee (C_a(z) \wedge \neg P_a(w)) \vee \neg w < y$. The only way to try to make $(**)$ fail at M_1 is to take x and z to be T or G, and to take both y and w to be 1. However, this valuation makes $\neg w < y$ to be the true formula $\neg 1 < 1$, making $(**)$ true at M_1. Thus *Axiom* (7)(a) is true at both M_0 and M_1, and similarly for *Axiom* (7)(b). Therefore *Axiom* (7) is true at both M_0 and M_1. A similar argument works for *Axiom* (8). The first conjunct of *Axiom* (8) is *Axiom* (8)(a): $(\forall x)(PC(x) \wedge K(\forall y)(\forall z)(\forall w)(\forall u)(PC(y) \wedge PC(w) \wedge ((C_a(x) \wedge C_b(y) \rightarrow P_a(z)) \wedge (C_a(w) \wedge C_b(y) \rightarrow P_a(u)) \wedge \neg x = w \rightarrow u < z)) \rightarrow C_a(x)))$. This is of the form $(\forall x)(PC(x) \wedge KA \rightarrow C_a(x))$,which is LPK equivalent to $(+)$: $(\forall x)(\neg PC(x) \vee \neg KA \vee C_a(x))$. When x takes any of the values $0, 1, 5$ or 6, $\neg PC(x)$, and hence $(+)$, is true at both M_0 and M_1. When $x = T$, $C_a(x)$, and hence $(+)$, is true at both M_0 and M_1. It remains to deal with case $x = G$. When $x = G$ the formula $\neg A$ is LPK equivalent to $(++)$: $(\exists y)(\exists z)(\exists w)(\exists u)(PC(y) \wedge PC(w) \wedge ((C_a(G) \wedge C_b(y) \rightarrow P_a(z)) \wedge (C_a(w) \wedge C_b(y) \rightarrow P_a(u)) \wedge \neg G = w \wedge \neg u < z)$. To make $(++)$ true at M_1 take $y = w = T$ and $z = u = 1$. Thus $\neg A$ holds at M_1 under this valuation, so that $\neg KA$ holds at both M_0 and M_1 under this valuation. Therefore $(+)$ also holds at M_0 and M_1 for $x = G$, and $(+)$ holds at both M_0 and M_1 for all values of x. Thus *Axiom* (8)(a) holds at both M_0 and M_1, and similarly for *Axiom* (8)(b). Therefore *Axiom* (8) is true at both M_0 and M_1. Thus all the axioms of (SGC) hold at the base world M_0. Since M_0 is a consistent LP world, for no extensional (i.e. K free) sentence A is it the case that both A and $\neg A$ hold at M_0. Since M is an LPK model of (SGC) it follows by the soundness meta-theorem for LPK that for no K-free formula A is it the case that $A \wedge \neg A$ is derivable in LPK from (SGC). $\square$

It is interesting to observe that it is the inconsistency of the world M_1 in the above model that is responsible for making both *Axioms* (7) and (8) true in the model. M_1 is inconsistent because both players choose both T and G in M_1. If one of the players were to choose only G in M_1 then that player would not choose the dominant strategy, and *Axiom* (8) would fail. If none of the players were to choose G, then the proof of *Axiom* (7) would break down at the following point; the formula $(**)$: $\neg PC(x) \vee \neg PC(z) \vee (C_a(x) \wedge \neg P_a(y)) \vee (C_a(z) \wedge \neg P_a(w)) \vee \neg w < y$ would fail in M_1 by taking $x = z = G$, $w = 1$ and $y = 5$.

Theorem 6. *There is a logical contradiction derivable from the Soros Game Context of the form* $KA \wedge \neg KA$.

Proof. Let A be a formula such that $(SGC) \vdash_{LPK} A \wedge \neg A$. By rule $(KR1(b))$ of LPK, $(SGC) \vdash_{LPK} A \wedge \neg KA$. By *Axiom* (9) of (SGC), and axiom $(A3)$ and rules $(KR2(a))$, (KMP) and (MRK) of LPK, $(SGC) \vdash_{LPK} KA \wedge \neg KA$. □

The main consequence of Theorems 5 and 6 for the rational self-interest assumption of economic theory is that in conflict-reflexive contexts the decision-making of rational self-interested agents is stymied by the appearance of true logical contradictions of the form $KA \wedge \neg KA$ ('Something is both known and not known').

7 Priest on backwards inductions

Priest (2000) proves the inconsistency of natural formalisations of the Surprise Examination Context, and of the Centipede Game Context of Rosenthal (1982). Priest's inconsistency results are summarised in Section 7.1. The resolution of these inconsistencies within the LPK framework of this paper is given in Section 7.2. Priest's response to the inconsistencies arising in game contexts is given in Section 7.3. The comparison of Priest's and the authors' interpretations concludes in Section 7.4.

7.1 Priest's formalisations of the Surprise Examination and Centipede Game Contexts

Priest (2000) gives a formalisation of the Surprise Examination Context, (SEC), making essential use of an intensional implication operator, $\rightarrow_P$, that is different from material implication, and which is detachable (i.e. Modus Ponens: $A, A \rightarrow_P B \Rightarrow B$, holds for $\rightarrow_P$). From (SEC) the logical contradiction $\alpha(1) \wedge \neg\alpha(1)$ is derived, where $\alpha(1) \wedge \neg\alpha(1)$ is read 'There both will and will not be a first examination on one of the days $1, ..., n$'. What makes this contradiction disturbing is that it is a contradiction

of the Beach Ball kind, i.e. (SEC) implies an inconsistency in the realm of concrete physical objects, since a first examination occurring and not occurring on one of the days is like a red-all-over and green-all-over beach ball.

Priest also presents a formalisation of the Centipede Game Context, again making use of a detachable intensional implication operator $\rightarrow_P$. In Priest (2000, Footnote 30) he gives a many-worlds model showing that his formalisation of the Centipede Game Context is *consistent*, and in Priest (2000, Footnote 31) he notes that his formalisation becomes *inconsistent* if $\rightarrow_P$ is taken to be material implication and if the obviously true axiom 'No player can take \$1 and \$2 on the same move' is added to the context (an axiom that for some reason he does not include in his main discussion of this context).

7.2 The LPK^n interpretation of Priest's results

LPK^n refers to the family of systems obtained from LP by the addition of modal operators K^n, where the intended interpretation of K^nA is 'A is known at time $t = n$'. Detailed specifications of the LPK^n systems are not required in order to outline the authors' interpretation of Priest's results.

The LPK^n treatment of the Surprise Examination Context is to take Priest's implication $\rightarrow_P$ to be material implication, as in the Soros Game Context. In that case Modus Ponens for $\rightarrow_P$ fails, and the derivation of the Beach Ball type inconsistency $\alpha(1) \wedge \neg\alpha(1)$, using LPK^n as base logic is no longer possible. The Surprise Examination Context remains inconsistent, but the inconsistency is now of the Soros Game Context kind, i.e. what are derivable are not concrete Beach Ball type inconsistencies, but only true contradictions essentially involving the knowledge operator K.

The authors' treatment of Priest's results for the Centipede Game Context is the same as for the Surprise Examination Context. The implication operator $\rightarrow_P$ is again interpreted as material implication, and LPK^n is used as base logic. Exactly as for the Soros Game and Surprise Examination Contexts, the Centipede Game Context is inconsistent, but no inconsistencies of the Beach Ball kind are derivable, as can be seen by suitably adapting the many-worlds model of Priest (2000, Footnote 30) and applying the relevant soundness result.

Thus, the LPK^n approach leads to a uniform treatment of all three contexts, and perhaps of inconsistent, conflict-reflexive contexts in general. It is important to note, however, that such formalisations, like the formalisation of the Soros Game Context in Section 6, are incomplete. The absence of a detachable intensional implication operator in LPK^n means that there are certain intensional facts about

these contexts that cannot be formalised in LPK^n. Note too though, that although these LPK^n formalisations are incomplete, they are sound (the axioms are true under the intended interpretation), and the formalisations are sufficiently complete for the logical contradictions to be derivable. Thus, the LPK^n framework allows some significant initial progress to be made in investigating the inconsistent structure of game contexts, despite the fact that the LPK^n formalisations are blind to those intensional properties of game contexts that can be captured only by a detachable intensional implication operator.

7.3 Priest's interpretation

Priest's response to the derivations of inconsistencies in game contexts differs from that of the authors. The main point of difference is as follows, and is discussed first in relation to the inconsistency results of the Soros Game Context.

Priest takes issue with the use of material implication in *Axioms* (7) and (8) of the Soros Game Context[1]. Only *Axiom* (7) is discussed, with similar remarks being applicable to *Axiom* (8). *Axiom* (7) formalises the concept of an ideally rational and self-interested agent, where an ideally rational and self-interested agent is one who actually behaves rationally and self-interestedly. Technically this idealness is formalised by taking $\rightarrow$ as material implication. When $\rightarrow$ is material implication, *Axiom* (7) says that if a's choosing x at time $t = 0$ gives more money at time $t = 1$ than choosing z at time $t = 0$, then a actually chooses x at time $t = 0$ (and similarly for b), i.e. a and b behave ideally rationally in the Soros Game because they actually do what is rational. Priest takes the proof of the inconsistency of the Soros Game Context in Theorem 4 as a *reductio* proof; from Theorem 4 he concludes that it is logically impossible for there to be such ideally rational and self-interested agents, just as the paradox of the Barber of Seville shows the existence of such a barber to be logically impossible.

This does not mean that he accepts the truth of the negation of *Axiom* (7). Rather, he takes *Axiom* (7) to be vacuously true. Putting in the universal quantifiers before *Axiom* (7) gives:

> For all agents a and b who behave ideally rationally and self-interestedly in the context of the Soros Game, *Axiom* (7).

For Priest it is logically impossible for there to be any such ideally rational and self-interested agents and so the axiom is vacuously true. In contrast, the authors'

[1]Personal communication.

view is that agents can, and in fact sometimes do, behave ideally rationally, particularly in certain simple, well-defined, closed contexts such as versions of the game contexts under consideration, and they therefore take *Axiom* (7) (and *Axiom* (8)) to be non-vacuously true for such contexts.

A detailed Priestian formalisation of the Soros Game Context using an intensional detachable implication, $\rightarrow_P$, is not presented here. For present purposes it is enough to say that such a formalisation, if it were sufficiently complete, would lead to rational dilemmas, i.e. to true sentences of the form: 'Agent a is rationally obliged to do x and at the same time agent a is rationally obliged not to do x'. Priest (2006(b), Chapter 6) derives such a rational dilemma in a Prisoners' Dilemma Context (of which the Soros Game Context is a special case) by formalising his non-ideal rationality principles and using essentially the same argument as in the proof of Theorem 4.

Making use of a modal rational obligation operator O, a rational dilemma is a true sentence of the form $OA \wedge O\neg A$. Note that this is not a true contradiction because it is not of the form $B \wedge \neg B$. The authors' formalisation of the Soros Game Context, on the other hand does lead to true contradictions of the form: 'A is known and at the same time it is not the case that A is known', i.e. to true contradictions of the form $KA \wedge \neg KA$. From the point of view of the many-worlds semantics of Section 5, and its various extensions, the difference between Priest's and the authors' interpretations is that in Priest's formulation true contradictions can occur only in possible worlds, but not in the actual world (the base world). If the rational dilemma $OA \wedge O\neg A$ holds in the actual world, then the contradiction $A \wedge \neg A$ holds in a possible world, while in the authors' formulation the contradiction $KA \wedge \neg KA$ holds in the actual world.

Thus, in the Soros Game Context Priest avoids the appearance of true contradictions holding in the actual world by denying the logical possibility of ideally rational behaviour.

Regarding the Centipede Game Context Priest does not need to do anything, since, as noted in Section 7.1 the formalisation of the Centipede Game Context using Priest's non-material $\rightarrow_P$is already consistent.

Priest does need to do something, however with his formalisation of the Surprise Examination, since even with the Priestian non-material $\rightarrow_P$ a true contradiction holding in the actual world is obtained. Priest's solution is to reject the principle of the persistence of knowledge over time, rule $(K) : K^iA \Rightarrow K^{i+1}A$, and if that is done the derivation of the inconsistency no longer goes through. The final sections of his paper are mainly concerned with arguing that it is plausible to reject (K) in the context of the Surprise Examination, and in other backwards induction contexts. In contrast the authors retain rule (K) in backwards induction contexts, their view being

that, while there are strong arguments for rejecting (K) in general, the persistence of knowledge over time is a highly plausible principle when applied to sufficiently simple, well-defined, closed contexts, including many versions of the Surprise Examination and Centipede Game Contexts. Retaining rule (K) ensures the derivation of true non-Beach Ball type contradictions holding in the actual world.

Thus, in the Surprise Examination Context Priest avoids the appearance of true contradictions holding in the actual world by denying the logical possibility of ideal knowledge that persists over time.

Overall, Priest's strategy in dealing with inconsistent game contexts is to argue that derivations of logical contradictions holding in the actual world should be interpreted as *reductio* arguments, proving the logical impossibility of certain idealisations of rationality and knowledge. The authors', in contrast, bite the other bullet, retaining the logical possibility of ideal rationality and knowledge at the cost of accepting the existence of true logical contradictions (though not of the Beach Ball kind) holding in the actual world.

7.4 Concluding remarks on Priest's interpretation

From the discussion of Section 7.3 an important point arises, namely, that from the point of view of the main thesis of this paper on the logical limits of rational self-interest as a foundation for economic theory, Priest's position is even more damaging. Under the authors' interpretation, in conflict-reflexive contexts the decision-making of ideally rational self-interested agents will be stymied by the appearance of true contradictions of the form $KA \wedge \neg KA$ (A is both known and not known). This is bad enough. However, under Priest's interpretation it is logically impossible for there to be any ideally rational self-interested agents to be making decisions in the first place. Further, under Priest's interpretation the decision-making of self-interested and non-ideally rational agents will be stymied by the appearance of rational dilemmas of the form $OA \wedge O\neg A$ ('I rationally ought to do A and I rationally ought not to do A').

8 Concluding remarks

In the paper it is shown that a natural formalisation of the Soros Game Context in first order logic is inconsistent, and interpreting the results of Priest (2000) within the paraconsistent framework shows the inconsistency of natural formalisations of the Surprise Examination and Centipede Game Contexts. The inconsistencies that arise always contain the knowledge operator K. No inconsistency can be proved

purely in the realm of concrete physical objects, i.e. the Beach Ball Problem does not arise.

The derivations of the inconsistencies in these contexts are sufficiently uniform and routine that it is conjectured that similar results are provable for a very wide class of game contexts which exhibit both conflict and reflexivity. Such conflict and reflexivity is apparent in many business, economic and other social contexts. Indeed, except for special limiting cases such as perfect competition, the existence of conflict and reflexivity in strategic interactions between economic agents appears to be the rule rather than the exception.

It should not, perhaps be surprising that conflict-reflexive social contexts are inconsistent. Such contexts are analogous to the following version of the Liar Paradox:

(α) Sentences α and β are not both true

(β) Sentences β and α are not both true

The Liar sentences exhibit conflict; (α) undermines (β), and vice versa. The Liar sentences also exhibit reflexivity; from the symmetrical nature of the Liar it follows that each sentence undermines, not only the other sentence but also itself. It is the combination of conflict and reflexivity that leads to the paradox.

In the Soros Game the players are in conflict with each other (from the self-interest assumption), and reflexivity follows from the symmetrical nature of the game. By symmetry, for every formula $A(a)$ true of a, $A(b)$ must also be true of b, and vice versa, i.e. a and b are indiscernibles. All this is common knowledge. Therefore, the rational self-interested players know that it follows from the meaning of the Soros Game Context that the choice made by each player is logically dependent on the choice made by the other player. Again, it is the combination of conflict and logical dependence that leads to paradox.

The above discussion of the logical dependence holding between the choices made by the players makes use of intensional notions. LPK, lacking a detachable intensional implication operator, is unable to state this logical dependence explicitly. However, sufficient of the reflexivity of the Soros Game Context is expressible implicitly in the language of LPK for the contradictions to be derivable.

Considering the Liar Paradox aspects of conflict-reflexive social contexts, it would perhaps be more surprising if such contexts were not inconsistent.

From the standpoint of paraconsistent mathematics, inconsistent conflict-reflexive social contexts exist (at least in the same sense that the counting numbers $1, 2, 3, ...$exist, in whatever sense that may be), and true logical contradictions, of the non-Beach Ball kind, hold of these contexts in the actual world. Inconsistent conflict-reflexive

situations arise when social agents engage in social interactions following the principles of rationality and self-interest. In that case the decision-making process is stymied by the appearance of true logical contradictions. Rational self-interested decision-making then becomes logically impossible.

However, this excludes only rational choice in combination with self-interest; rational decision-making is possible provided such decision-making occurs within an appropriate moral framework. Within the constraints given by such a moral framework, rational decision-making is no longer stymied by the appearance of inconsistencies, and rational choice then becomes possible. For example, following the Golden Rule of treating the other player in the way one would like to be treated oneself, perhaps the most widely-accepted conventional moral principle, leads the players in the Soros Game each to give \$5, and for the players in the Centipede Game to take turns and share the money equally (In the examples given in the paper, our fictional friends Allie and Bobbie do in fact make the rational moral choice, each giving the other player \$5). In the Soros Game Context, treating the decision as a rational moral choice, rather than as a choice based on rational self-interest, amounts to deleting the rational self-interest axioms, *Axiom* (7) and *Axiom* (8) of the Soros Game Context, and replacing them with an axiom stating that each player makes the choice that maximises the payoff to the other player, given that the other player follows the same altruistic strategy. In this case no inconsistency arises. Similarly, the conventional moral framework leads to consistent formalisations of the Centipede Game Context and of many other conflict-reflexive social contexts.

Acknoweldgements

The authors would like to thank Graham Priest and Barry Murphy for their insightful criticisms of an earlier draft of this paper. We would also like to express our thanks to the anonymous referee, whose comments lead to a number of improvements made in the final version.

References

Daynes, A., 2000. "A strictly finitary non-triviality proof for a paraconsistent system of set theory deductively equivalent to classical ZFC minus foundation" *Archive for Mathematical Logic* 39, 581-598.

Gintis, H., 2000. *Game Theory Evolving: A Problem-Centred Introduction to modelling Strategic Interaction*, Princeton University Press, Princeton, NJ.

Priest, G., 1979. "Logic of paradox" *Journal of Philosophical Logic* 8, 219-241.

Priest, G., 2000. "The logic of backwards inductions" *Journal of Economics and Philosophy* 16, 276-285.

Priest, G., 2006(a). *In Contradiction*, 2nd edition. Oxford University Press, Oxford.

Priest, G., 2006(b). *Doubt Truth to be a Liar*, Oxford University Press, Oxford.

Priest, G., 2008. *An Introduction to Non-Classical Logic: From If to Is*, 2nd edition. Cambridge University Press, Cambridge.

Rosenthal, R., 1982. "Games of perfect information, predatory pricing and the chain-store paradox" *Journal of Economic Theory* 25, 92-100.

Soros, G., 1987. *The Alchemy of Finance: Reading the Mind of the Market*, Simon and Shuster, New York.