Cooperation and defection in a generalized modal setting

T. Nicholson
Laboratory for Logic and Experimental Philosophy
Department of Philosophy
Simon Fraser University
Burnaby, BC, Canada
V5A 1S6

Abstract *A modal logical schema is introduced for the exploration of a multi-player generalization of the Prisoner's Dilemma in which (a) each participant has at least k available moves, and (b) participants can be members of more than one coalition of successful participants. The methodology employed illustrates how the principle governing the aggregative behaviour of formulae within the scope of the* $\Box$ *operator in the models for a class of non-standard modal logics can be manipulated to represent list-colouring properties for hypergraphs.*

Keywords: modal logic, hypergraphs, list-colourings, Prisoner's Dilemma, $(n+1)$-ary relational model

Research supported by Natural Sciences and Engineering Research Council of Canada grant 232471-2000

1 Introduction

In [3], [5], and [6], a key relationship is demonstrated between the semantic theory for a class of weakly aggregative modal logics, and the theory of hypergraph colouring. In this article, this relationship will be further exploited in the development of a multi-player generalization of the Prisoner's Dilemma which is intended to model some of the complexities of player interactions arising from the formation of coalitions among proper subsets of participants.

2 Hypergraphs and Modal Aggregation

An $(n+1)$-*ary relational model* $\mathcal{M}$ is a triple $\langle U, R, \mathcal{V}\rangle$, where U is a non-empty set of points, $R \subseteq U^{n+1}$, and $\mathcal{V}$ is a valuation function from a denumerable set of atoms to $\wp(U)$, which is defined standardly for Boolean connectives, and as follows for $\Box$:

for every sentence α, for every $x \in U$,

$$x \in \mathcal{V}(\Box\alpha) \Leftrightarrow$$

$$\forall\langle y_1, y_2, ..., y_i, ...y_n\rangle \in U^n,$$

$$Rxy_1, ..., y_n \Rightarrow$$

$$\exists i \ (1 \leq i \leq n) : y_i \in \mathcal{V}(\alpha).$$

The phrase "weakly aggregative" has been used to describe the logics of these models because the strong aggregation principle

$$K : \Box p \wedge \Box q \to \Box(p \wedge q)$$

is not valid, and is generalized by a weaker formulation involving hypergraphs of chromatic number strictly greater than n.

Definition 2.1 *Let V_H be a finite non-empty set. Let $\mathcal{E}_H \subseteq 2^V - \{\emptyset\}$. Then $H = \langle V_H, \mathcal{E}_H \rangle$ is a* hypergraph*, with V_H the set of H-*nodes*, and $\mathcal{E}_H$ the set of H-*edges[1]*.*

Intuitively, the chromatic number of a hypergraph H refers to the least number of parts into which its nodes must be divided to prevent any edge of H from appearing intact in any of the parts. Further, any such division of H-nodes is called a (proper) colouring of H. More formally:

Definition 2.2 *Let H be a hypergraph. Then a* (proper) k-colouring *of H is a family $C = \{c_i\ (1 \leq i \leq k) \mid \bigcup_{i=1}^{k} c_i = V, \forall i, j, i \neq j \Rightarrow c_i \cap c_j = \emptyset$, and $\forall e \in \mathcal{E},\ e \not\subseteq c_i,\ (1 \leq i \leq k)\}$.*

Definition 2.3 *Let H be a hypergraph. Then the* chromatic number *of H , $\chi(H)$, is the least integer k for which there is a k-colouring of H.*

Definition 2.4 *Let H be a hypergraph whose nodes are sentences[2]. Then the* formulation *of H, $F(H)$ is the following sentence:*

$$\bigvee_{i=1}^{|\mathcal{E}|} \bigwedge_{h=1}^{|e_i|} h \in e_i$$

Now from the definition of $\mathcal{V}$ for $\Box$ on an $(n+1)$-ary relational model $\mathcal{M} = \langle U, R, \mathcal{V} \rangle$, for each point $x \in U$, it follows that every $\langle y_1, y_2, ..., y_n \rangle \in U^n$ such that $\langle x, y_1, ..., y_n \rangle \in R$ induces an n-decomposition of $\Box(x) = \{\beta \mid x \in \mathcal{V}(\Box\beta)\}$ [6], viz., $\{d_i\ (1 \leq i \leq n) \mid d_i = \{\beta \in \Box(x) \mid y_i \in \mathcal{V}(\beta)\}\}$. Consequently:

Theorem 2.5 *For every $(n+1)$-ary relational model, for every hypergraph $H = \langle [j], \mathcal{E} \rangle$ whose nodes are atoms, $\chi(H) > n \iff$ the following sentence is valid* [4],[6]*:*

$$\Box 1 \wedge \Box 2 \wedge ... \wedge \Box j \rightarrow \Box F(H)$$

However, it is not so much the colouring properties of hypergraphs which are interesting from the point of view of the game-theoretic application at hand, but rather the generalized notion of list-colouring properties.

Definition 2.6 *Let j, k, and n be positive integers ($n \geq k$). Then where $H = \langle [j], \mathcal{E} \rangle$ is a hypergraph, a k-*list assignment μ in $[n]$ *to H is a function $\mu : [j] \rightarrow 2^n$ where $\forall i \in [j], |\mu(i)| \geq k$.*

Definition 2.7 *Let $H = \langle [j], \mathcal{E} \rangle$ be a hypergraph, with k and n positive integers ($k \leq n$), and μ a k-list assignment in $[n]$. Then a function λ is* a (proper) colouring of H on μ *(the function λ (properly) colours H on μ) if $\lambda : [j] \rightarrow [n]$ satisfying:*

1. *$\forall i \in [j]$, $\lambda(i) \in \mu(i)$, and*
2. *$\forall e = \{1, 2, ..., j\} \in H$, $\exists h, j \in e : \lambda(h) \neq \lambda(j)$.*

By exploiting this notion of list colouring for hypergraphs, in what follows a generalization of the Prisoner's Dilemma is given, a decision problem of which is reduced to satisfiability on $(n+1)$-ary relational models.

3 Standard form PDs

In its standard form the Prisoner's Dilemma (PD) can be represented as a k-player game. Where $M = \{C, D\}$ is the set of available moves, ("C" representing "cooperate", and "D" representing "defect"), and $\forall i (1 \leq i \leq k), \forall X = \langle 1, 2, ..., i, ..., k \rangle \in M^k$, $X_i(C) := \langle 1, 2, ..., C, ..., k \rangle$, and $X_i(D) := \langle 1, 2, ..., D, ..., k \rangle$, a score assignment function S_i is defined for each player P_i ($1 \leq i \leq k$), arbitrary in its particulars, but necessarily satisfying the following conditions:

[1] Following convention (e.g., see [1]), "H" will often be used to denote $\mathcal{E}$. Essentially this is because the present context renders so-called "isolated nodes" (those not occurring in edges) relatively uninteresting. For similar reasons, also following convention, attention will be restricted to hypergraphs without singleton edges. Lastly, subscripts will be dropped for perspicuity, when context allows.

[2] Any standard modal language with denumerably many atoms will do.

1. $S_i : M^k \rightarrow N$,

2. $S_i(\{D\}^k) < S_i(\{C\}^k)$, and

3. $S_i(X_i(D)) > S_i(X_i(C))$.

These score assignment functions represent the essence of the PD, and hence that its dilemmatic character is often expressed as a conundrum of practical rationality. For, apparently, rationality occasionally prescribes (putatively) irrational behaviour. In this case, e.g., it seems as though each PD player ought to defect, notwithstanding that unanimous cooperation yields a better return than when all players defect.

One reason to generalize the Prisoner's Dilemma (PD), aside from the usual interest in applications to economic, political, social and biological phenomena, has to do with a failure of standard formulation PDs to accommodate the following facts:

1. Participants often have several available moves.

2. The moves available to one participant may not be the same as the moves available to another.

3. Participants often belong to several distinct coalitions of participants.

4. The effect of cooperation among the members of a coalition, with respect to the final outcome of the interaction, can differ from coalition to coalition.

While it is difficult to say exactly why standard form PDs are limited in these ways, one contributing factor seems to lie in the interface between its formalization and its intended interpretation(s). Specifically, because cooperation between participants is modeled by the mutual selection of the move "cooperate" among players, rather than some richer set of structural relations between available moves, move selections, and players, once additional moves are made available to participants, exactly what cooperation among players amounts to becomes something of an open question. To remedy this, in what follows, *cooperation between players will be said to occur exactly when each player selects the same move*. Evidently, this use of "cooperate" is different than that in PDs, since the players of those games are not to be seen as cooperating when each chooses "defect". However, it is arguable that this difference is nominal, insofar as a unanimous choice among PD players to defect involves a decision on the part of player 1 to defect *from player 2*, for player 2 to *defect from player 1*, etc., and therefore does not reflect an identity of move selections at all.

Further, to make room for a representation of the possibly differing effects of cooperation among members of subsets of participants, we stipulate that *coalitions* among players represent subsets $e_1, e_2, ..., e_j$ of players satisfying the following condition: mutual cooperation among the members of $e_i (1 \leq i \leq j)$ is more beneficial than if one of them succumbs to a (rational) temptation to not cooperate with the group. Although a substantive interpretation of this use of "coalition" will not be given here, the idea is to assume that some such account has been given, in order to focus on the complexities of play arising from structural relations between moves available to players, actual move selections, and coalitions. Thus, in what follows, a coalition is said to be *successful* on the condition that its members cooperate (select the same move). For even in such liberally defined circumstances as these, interesting questions about the *efficacy of coalitions*, stemming from structural relationships between existing coalitions and the possible move selections made by participants may be asked, among them:

Is there a selection of moves by the participants such that no coalition succeeds?

Or, alternatively,

Is the case that no matter which moves participants select, there is always some successful coalition?

Presumably, interactions characterized by an affirmative answer to the latter question are to some extent weighted in favor of the survival, in the long run, of populations decomposable into the game's coalitions.

4 A Modal Reduction

Consider a game $\mathcal{G}$ involving j players, where the players $1, 2, ..., j$ are grouped into coalitions $e_1, e_2, ..., e_k$. This structure can be modelled using the hypergraph $\mathsf{H} = \langle [j], \{e_1, e_2, ..., e_k\}\rangle$. Accordingly, for any such game $\mathcal{G}$, we can say that $\mathsf{H}_{\mathcal{G}} = \langle [j], \{e_1, e_2, ..., e_k\}\rangle$ is the *coalition hypergraph* for $\mathcal{G}$. Further, where k is the least number of moves available to any player in the game, and $[n]$ is the union of the moves available to all players, let $\mu : [j] \to \wp([n])$, where for each player $i \in [j]$, $\mu(i)$ is the set of moves available to i. Then μ is a k-list assignment in $[n]$ for the coalition hypergraph $\mathsf{H}_{\mathcal{G}}$, and:

Theorem 4.1 *There is a function $\lambda : [j] \to [n]$ which properly colours $\mathsf{H}_{\mathcal{G}}$ on μ iff the players of $\mathcal{G}$ may individually select moves in such a way that no coalition succeeds.*

At this point we are in a position to show, relative to a fixed k-list assignment μ in $[n]$ for the coalition hypergraph $\mathsf{H}_{\mathcal{G}}$, that the problem of determining whether or not the players of $\mathcal{G}$ can select moves in such a way that no coalition is successful is reducible to the satisfiability problem for $(n+1)$-ary relational models. In order to do this, it suffices to massage the antecedent of the formula in Theorem 2.5 in such a way to force any model $\mathcal{M} = \langle U, R, \mathcal{V}\rangle$ on which the negation of the resulting sentence is true to satisfy the following condition: for some $x \in U$, there is a related n-tuple $y_1, y_2, ..., y_n$ which induces an n-decomposition of $\Box(x)$ from which we may obtain a proper colouring λ on μ, while at the same time having no point y_j $(1 \leq j \leq n)$ in the set $\mathcal{V}(\mathsf{F}(\mathsf{H}))$. Thus, among other things, we need atoms which represent players, in addition to atoms which represent moves, and a player must be forced to select only from among those moves which appear in the subset of $[n]$ determined for her or him by the list-assignment μ. Further, we need some way of representing move selections by players, and perhaps most importantly, we need to ensure that any subset of players all of whose members select the same move is such that its members all appear, at least once, in the same cell of the n-decomposition induced by $y_1, y_2, ..., y_n$. Essentially this is because, in effect, a colouring λ on μ in $[n]$ will be extracted from this n-decomposition by simply deleting elements of the decomposition's cells.

To this end, let μ be a k-list assignment in $[n]$ for the hypergraph $\mathsf{H}_{\mathcal{G}} = \langle [j], \mathcal{E}\rangle$. Let $\{p_1, p_2, ..., q_1, q_2, ...\}$ be a denumerable set of atoms, where p_i represents player i, and q_m represents move m. Then

Definition 4.2 *The* move-selection sentence *for H on μ, M_f^{H} is the sentence:*

$$\bigwedge_{i=1}^{j} \bigvee_{m \in \mu(i)} \Box(p_i \wedge q_m)$$

The idea is that if M_μ^{H} is satisfied on a model $\mathcal{M} = \langle U, R, \mathcal{V}\rangle$, then at some $x \in U$, for each player we have the $\Box$ of the conjunction of the player's atom with the atom corresponding to at least one of his or her available moves. At this point, however, from the definition of the valuation function $\mathcal{V}$ for $\Box$, it does not follow that x is related to any n-tuple $\langle y_1, y_2, ..., y_n\rangle \in U^n$ such that for each pair p_h, p_i of distinct player atoms, if h and i end up selecting the same move then for some f $(1 \leq f \leq n)$, $y_f \in \mathcal{V}(p_h) \cap \mathcal{V}(p_i)$. But to force this, we can manipulate the "shepharding" behaviour of $\Box$ on these models in the following way:

Definition 4.3 *Let $\mathsf{H}_{\mathcal{G}} = \langle [j], \mathcal{E}\rangle$ be the coalition hypergraph for $\mathcal{G}$, let μ be a k-list assignment in $[n]$, and let $\{A_1, A_2, ..., A_r\}$ be the set* *of all subsets of [j] of size at least 2. Then where P_{A_i} is the set of atoms corresponding to the elements of A_i $(1 \leq i \leq r)$,*

the aggregation sentence *for* H *on* μ, $A_\mu^{\mathcal{G}}$, *is the sentence*

$$\bigwedge_{m=1}^{n}(\bigwedge_{k=1}^{r}((\bigwedge_{i=1}^{|A_k|}\Box(p_{i\in A_k}\wedge q_m))\rightarrow\Box(\wedge[P_{A_k}]\wedge q_m)))$$

Thus, if $M_\mu^{\mathsf{H}} \wedge A_f^{\mathsf{H}}$ is satisfied on a model $\mathcal{M} = \langle U, R, \mathcal{V}\rangle$, at some $x \in U$, not only do we have a representation of each player selecting an appropriate move (as determined by μ), but all players selecting the same move are aggregated with that move in the following sense:

$\forall\langle y_1, y_2, ..., y_n\rangle \in U^n : Rxy_1, ..., y_n, \forall A_i \subseteq [j]$ $(|A_i| \geq 2)$, if every member of A_i selects the same move m, then for some f $(1 \leq f \leq n)$, every atom in P_{A_i} is true at y_f, and so is q_m.

Therefore, where the *game sentence*, $\mathcal{G}_\mu^{\mathsf{H}}$ for $\mathsf{H}_\mu^{\mathcal{G}}$ is the sentence:

$$M_\mu^{\mathsf{H}} \wedge A_\mu^{\mathsf{H}} \rightarrow \Box\mathsf{F}(\mathsf{H}),$$

it follows that:

Theorem 4.4 *The players of $\mathcal{G}$ may individually select moves in such a way that no coalition succeeds iff $\neg\mathcal{G}_\mu^{\mathsf{H}}$ is satisfiable on an $(n+1)$-ary model.*

Proof.

[⇒]

By Theorem 4.1 we may begin by assuming that there is a proper colouring λ on μ of $\mathsf{H}_\mu^{\mathcal{G}}$. Then there is a family of sets $C = \{c_i$ $(1 \leq i \leq n) \mid \bigcup_{i=1}^{n} c_i = \mathsf{V}$ and $\forall i, j, i \neq j \Rightarrow c_i \cap c_j = \emptyset\}$ where $\forall e \in \mathsf{H}, e \not\subseteq c_i$ $(1 \leq i \leq n)$, and for each $h \in [j]$, $h \in c_i$ only if $i \in \mu(h)$.

Construct a model $M = \langle U, R, \mathcal{V}\rangle$ as follows:

1. set $U := \{x, c_1, c_2, ..., c_n\}$,
2. set $R \subseteq U^{n+1} := \{\langle x, c_1, c_2, ..., c_n\rangle\}$,
3. for every atom p_j, set $\mathcal{V}(p_j) := \{c_i\}$ if $j \in c_i$; else set $\mathcal{V}(p_j) := \emptyset$, and
4. for every atom q_h, set $\mathcal{V}(q_h) := \{c_i\}$ if $h = i$; else set $\mathcal{V}(q_h) := \emptyset$.

Then since no H-edge is a subset of any element of C, it follows that $x \in \mathcal{V}(\neg\mathsf{F}(\mathsf{H}))$. Further, since every player has selected an appropriate move in the game $\mathcal{G}$, it follows that (1) for every atom corresponding to a player, the $\Box$ of the conjunction of that atom with at least one atom representing an appropriate move for the player is true at x, and (2) for every subset A_i of players all selecting the same move m, the $\Box$ of the conjunction of their corresponding atoms with q_m is true at x.

Therefore $\neg\mathcal{G}_\mu^{\mathsf{H}}$ is true at x in $\mathcal{M}$.

[⇐]

Now suppose that for some model $\mathcal{M} = \langle U, R, \mathcal{V}\rangle$, for some $x \in U, x \notin \mathcal{V}(\neg\mathcal{G}_\mu^{\mathsf{H}})$. Then by the definition $\mathcal{V}$, the antecedent of $\mathcal{G}_\mu^{\mathsf{H}}$ holds at x, and the consequent fails. Therefore, there is an x-related n-tuple $\langle y_1, y_2, ..., y_i, ..., y_n\rangle \in U^n$ such that for each i $(1 \leq i \leq n)$, for each $e \in \mathsf{H}$, not all atoms corresponding to nodes in e are true at y_i (else $\Box\mathsf{F}(\mathsf{H})$ is true at x, contrary to hypothesis), and, since $M_\mu^{\mathsf{H}} \wedge A_f^{\mathsf{H}}$ is true at x, for each $h \in [j]$ we have that:

1. $x \in V(\Box(p_h \wedge q_i))$, for some $i \in \mu(h)$, and
2. for every set A_j of H-vertices, if $\exists m \in [n] : \forall i \in A_j, x \in \mathcal{V}(\Box(p_i \wedge q_m))$, then $x \in \mathcal{V}(\Box \wedge [P_{A_j}] \wedge q_m)$.

Consequently, the n-decomposition $D = \{d_i$ $(1 \leq i \leq n) \mid d_i = \{\beta \mid y_i \in \mathcal{V}(\beta)\}\}$ induced by $\langle y_1, y_2, ..., y_i, ..., y_n\rangle$ may be used to define a proper colouring λ of H on μ:

First define a function $\lambda' : [j] \to 2^n$ such that $\forall h \in [j]$, $\forall i \in [n]$, $i \in \lambda'(h)$ iff

1. $i \in \mu(h)$, and
2. for some $d \in D$, $\{f \in [j] \mid i \in \mu(f)$ and $x \in \Box(p_f \wedge q_i)\} \subseteq d$.

Then a colouring $\lambda : [j] \to [n]$ may be defined from λ' by selecting a unique $i \in \lambda'(h)$, for each $h \in [j]$.

$\Box$

5 Conclusion

A natural extension of this research would be to augment the relational structures of $(n+1)$-ary relational models, or to introduce an alternative modal operator, with an eye towards exploring generalizations of chromatic number still further. For example, one could try to characterize the n-list chromatic number of an arbitrary hypergraph using $(n+1)$-ary relational models:

Definition 5.1 *Let H be a hypergraph. Then the n-*list chromatic number *of H is the least integer k ($k \leq n$) for which every k-list assignment μ in n to H admits of a proper colouring λ.*

Or, one could introduce the modal operator $\bigcirc$ for which the valuation function $\mathcal{V}$ on $(n+1)$-ary relational models is point-wise defined as follows:

$$\forall x \in U, x \in V(\bigcirc\alpha) \text{ iff}$$

$$\forall Y = \langle y_1, y_2, ..., y_n\rangle \in U^n, Rxy_1...y_n \Rightarrow$$

$$\exists Z = \langle z_1, z_2, ...z_i, ..., z_k\rangle \in Y^k \ (k \leq n):$$

$$\forall i \ (1 \leq i \leq k), z_i \in \mathcal{V}(\alpha).$$

For in that case it would seem as though something like the following result might be correct:

$\forall H = \langle [j], \mathcal{E}\rangle$, if H is a coalition hypergraph, then

$$\models \bigcirc p_1 \wedge \bigcirc p_2 \wedge ... \wedge \bigcirc p_j \to \Box F(H)$$

iff for every list assignment of available moves to players, there is at least one set of move selections by players on which some coalition is successful.

The reason why would have to do with the $\bigcirc$ operator's inverse representation of list-assignments: rather than each node in a coalition hypergraph being assigned some set of available moves, in effect each of the n moves available to any player is assigned to the set of players that would have that move available to them on the corresponding standard list assignment function.

References

[1] Berge, C. *Hypergraphs: combinatorics of finite sets*, North-Holland, 1989.

[2] Bollobás, B. *Modern Graph Theory.* Springer. New York. 1991.

[3] Jennings, R.E. and T. Nicholson, *A modal logical characterization of n-uncolourability for hypergraphs*, unpublished (2002).

[4] Jennings, R.E. P.K. Schotch. On Detonating. *Paraconsistent Logic.* (G. Priest and R. Routley, editors), *Philosophia* Verlag. Munich. 1989. pp.306-327.

[5] Nicholson, T. *A weakening of chromatic number*, Proceedings of the Student Session of the 13th European Summer School in Logic, Language, and Information (2001), 215–226.

[6] Nicholson, T., R.E. Jennings D. Sarenac. Revisiting completeness for the K_n modal logics: a new proof. *Logic journal of the IGPL*, vol. 8. 1, 2000.