The Big Picture

More Games of No Chance
MSRI Publications
Volume **42**, 2002

Idempotents Among Partisan Games

ELWYN BERLEKAMP

Abstract. We investigate some interesting extensions of the group of traditional games, $\mathcal{G}$, to a bigger semi-group, $\mathcal{S}$, generated by some new elements which are idempotents in the sense that each of them satisfies the equation $G + G = G$. We present an addition table for these idempotents, which include the 25-year-old "remote star" and the recent "enriched environments". Adding an appropriate idempotent into a sum of traditional games can often annihilate the less essential features of a position, and thus simplify the analysis by allowing one to focus on more important attributes.

1. Introduction and Background

We assume the reader is familiar with the first volume of *Winning Ways* [Berlekamp et al. 1982], including Conway's axiomatization of $\mathcal{G}$, the group of partesan games under addition, which can also be found in [Conway 1976]. I now call the elements of this group *traditional games*. Each of the traditional games considered in this paper has only a finite number of positions. The identity of $\mathcal{G}$ is the game called 0, which is an immediate win for the second player. We investigate some interesting extensions of the group $\mathcal{G}$ to a bigger semi-group, $\mathcal{S}$, generated by some new elements which are idempotents in the sense that each of them satisfies the equation $G + G = G$. We also present an addition table for these idempotents.

Some of these idempotents have long been well-known in other contexts. The newer ones all fall into a class I have been calling *enriched environments*. A companion paper [Berlekamp 2002] shows how these idempotents prove useful in solving a particular hard problem involving a gallimaufry of checkers, chess, domineering and Go.

We begin with a review of definitions, modified slightly to fit our present purposes.

Moves. In Go, a *move* is the change on the board resulting from the act of a single player. In chess, this is commonly called a *ply*, and the term *move* is used to describe a consecutive pair of plys, one by White and one by Black. In this paper, we use *move* as it is understood in Go. This is consistent with the tradition of combinatorial game theory. This theory has been most successful in analyzing positions which can be treated as sums of subpositions which are relatively or completely independent of each other. Each of these subpositions, as well as their sum, is called a *game*. Although the players alternate turns, it is quite common that they may elect to play in different components, so that within any particular game, the same player may make several consecutive moves. Hence, the definition of a game, or any of its positions, *does not* include any specification of whose turn it is to play next. Conway's traditional definition of a game is written recursively as

$$G = \{G^L \mid G^R\},$$

where G^L and G^R are sets of previously defined games. G^L is the set of positions to which Left can move immediately. These are also known as Left followers. G^R is the set of Right followers.

Other axioms, with which the reader is assumed to be familiar, define *sum*, *negative*, *greater-equal*, and *number*.

Outcomes. When played out, traditional games eventually yield outcomes. There are two outcomes: Loutcome, which is the outcome if Left plays first, and Routcome, which is the outcome if Right plays first. In the most general case, either of these outcomes might assume any of the values LEFT, TIE, or RIGHT, which Left prefers in the order LEFT > TIE > RIGHT. Ties and draws are impossible within $\mathcal{G}$, but they can occur in some of the extensions we will consider. In $\mathcal{G}$, play eventually terminates when the player to move is unable or unwilling to do so. That happens to Right when the value of the position is a nonnegative integer, or to Left when the value of the position is a nonpositive integer. If the value of the position is 0, whichever player is next to move is the loser.

Left plays to attain an outcome he prefers, while Right tries to thwart it. Loutcome and Routcome are the results if both players play optimally. If both

$$\text{Loutcome}(G) \geq \text{Loutcome}(H) \quad \text{and} \quad \text{Routcome}(G) \geq \text{Routcome}(H),$$

we say that

$$\text{Outcomes}(G) \geq \text{Outcomes}(H).$$

Greater-Equal. Combinatorial games satisfy a partial order. One form of the traditional definition of *greater-equal* states that

$$G \geq H \iff \text{For all } X,\ \text{Outcomes}(G+X) \geq \text{Outcomes}(H+X) \tag{1–1}$$

If H and X have negatives, this is equivalent to the assertion that

$$\text{Outcomes}(G-H) \geq \text{Outcomes}(0)$$

But since Loutcome(0) = RIGHT, that half of the condition is trivially satisfied so a sufficient condition is that

$$\text{Routcome}(G - H) = \text{LEFT},$$

or, as more commonly stated, Left, playing second, can win on $G - H$.

Following Conway's original axioms, we say that

$$G = H \iff G \geq H \text{ and } H \geq G$$

and that

$$G > H \iff G \geq H \text{ but } H \not\geq G$$

Scores. For some purposes, it is convenient to define *scores* and work with them rather than with outcomes.

Play of any traditional combinatorial game must eventually yield a position whose value is a number. The value of the first such position is called the game's *score* or *stop*. If Left plays first and G is optimally played, the resulting number is called the Leftscore, denoted by Lscore(G). Similarly, if Right plays first and G is optimally played, the resulting number is Rscore(G). If G is any traditional game and x is any number, then

$$x > \text{Lscore}(G) \Rightarrow x > G,$$
$$\text{Lscore}(G) > x > \text{Rscore}(G) \Rightarrow x \text{ is confused with } G,$$
$$\text{Rscore}(G) > x \Rightarrow G > x.$$

It is known that if G is any game, then an optimal Left strategy for playing G ensures reaching a maximal score, and an optimal Right strategy for playing G ensures reaching a minimal score. However, the converse need not be true because several strategies might lead to the same score and some of them might yield a suboptimal outcome. This is due to the fact that when the score of a traditional game is 0, the outcome depends on who gets the last move.

Some real games have other scores, which are explicitly defined by the rules of the game. Go is such a game. It is an initially surprising and somewhat remarkable fact that these *official scores* imposed by any of several variations of the *official* Go rules are often identical to these *mathematical scores*. By appropriate choices of rules for "Mathematical Go" [Berlekamp and Wolfe 1994], we can attain agreement of scores in all but a few very rare positions, which are so exotic that different variations of the official Go rules then fail to agree with each other.

Dots-and-Boxes is another popular game in which scores are explicitly defined by the rules of the game. This pencil-and-paper game has very little in common with Go. But surprisingly, it again happens that the elegant mathematics of combinatorial game theory, when applied to an approximation of the popular game, yields decisive insights into how to play the popular game [Berlekamp 2000b].

In this paper, we treat *scores* in the mathematical sense: the value of the first position whose value is a number.

We next consider several idempotent games that have no negatives.

2. Definitions of Idempotents with Opening Ceremonies

Remote Star. The remote star ☆ is introduced in *Winning Ways*, Chapter 8 and plays a leading role there. Rather than rely on any of those results for a definition, I now propose a rule for playing

$$Y + ☆$$

Before moving on such a game, we require an *opening ceremony* during which each player submits a positive integer to the referee as a sealed bid. (From the mathematical perspective, there is no need for these bids to be sealed; public bids would work equally well. However, professionals and other serious competitors are loath to let the opponent know anything about their contingency plans prematurely. So it is easier to sell mathematical models to serious game-players when the rules ensure that losing bids remain unknown to the winning bidder.)

The referee selects the larger bid, n, and replaces ☆ by $*n$ before play begins. Ordering relations are determined in the usual way, using (1–1), with the understanding that the game X is specified before the opening ceremonies.

The play of $Y + ☆ + ☆$ begins with two successive auctions. Since either player can submit a bid to the second auction which is at least double the winning bid of the first auction, it follows that the sum of two remote stars is now again a remote star, whence

$$☆ + ☆ = ☆.$$

Ish. Traditionally, the term *ish* means *Infinitesimally SHifted*. It appears in such expressions as $\{1 \mid 1\} = 1* = 1\,\text{ish}$ and $\{1* \mid 1\} = 1\,\text{ish}$.

We can also manipulate ish as though it were another element of our semigroup $\mathcal{S}$. To this end, we henceforth treat ish as a noun with the mnemonic Infinitesimal SHift. Its ordering relations might be defined as

$$G \le H + \text{ish} \ \text{ and/or } G + \text{ish} \le H \iff \text{Scores}(G+X) \le \text{Scores}(H+X) \text{ for all } X,$$

However, to simplify the task of defining the sum of ish plus other idempotents, I prefer the following more intricate definition of ish:

At the opening ceremony of $G + X + \text{ish}$, each player submits a small positive number as a sealed bid to the referee, who announces the smaller such bid, which we will call ε. Then Left wins only if the score exceeds ε, and Right wins only if the score is less than $-\varepsilon$. A game which concludes with a net score of magnitude not exceeding ε declared to be a tie.

It is not hard to show that Left, going first, is able to *win* the game $G+X+\text{ish}$ if and only if $\text{Lscore}(G + X) > 0$. Left, going second, is able to *win* the game

$G + X + \text{ish}$ if and only if $\text{Rscore}(G + X) > 0$. Furthermore, it is easily verified that

$$\text{ish} + \text{ish} = \text{ish}\,.$$

Comment. *The sophisticated reader will recall several types of numbers that appear in [Conway 1976]: surreal, real, rational, dyadic rational. So when reading that each bid to determine ε must be a small positive number, she might ask which sort of number is required. It turns out that any type of number just listed is adequate for our present purposes. But only dyadic rationals are fully consistent with our focus on games with a finite number of positions.*

Positively Enriched Environment, $\mathcal{E}_t$. Enriched environments entail more elaborate opening ceremonies. Each positively enriched environment has a specified temperature, t, which is a positive number.

Every enriched environment contains an implicit ish, which is resolved first. This results in the specification of an ε. If other ishes are present, then this initial portion of the opening ceremonies continues until are all resolved into a single small positive ε. Then, to resolve the positively enriched environment $\mathcal{E}_t$, each player submits to the referee another small positive number as a sealed bid. The referee announces the winning (smaller) number, called δ. For simplicity, we restrict legal bids so as to ensure that δ is a divisor of t, so that t/δ is a positive integer. Then $\mathcal{E}_t$ is replaced by the sum of t/δ uniformly spaced switches, called *coupons*: $t|{-t}$, $(t{-}\delta)|({-t}{+}\delta)$, $\ldots$, $\delta|{-\delta}$. This concludes the opening ceremonies. Then play begins. Play terminates when all coupons have been taken and the value of the position is a number. The net score is then taken to be this number plus all of Left's coupons minus all of Right's coupons. The outcome is declared to be a tie unless the magnitude of the score exceeds ε.

Fully Enriched Environment, $\mathscr{E}_t$. The temperature of a fully enriched environment can be any number not less than -1.

After resolving the implied and explicit ishes to an ε, each player submits a small positive bid for δ. Legal bids are constrainted to ensure that $(t + 1)/\delta$ is a nonnegative integer. The winning (smaller) bid is announced. Then $\mathscr{E}_t$ is replaced by a set of *coupons*, whose face values range from t down to $-1 + \delta$, with a constant decrement of δ. A very large number of coupons with value -1 is then placed at the bottom of the coupon stack.[1]

Play begins. At each turn, a player may either make a legal move from the current game, G, to one of his legal followers, *or* he may instead use his turn to

[1]Mathematically, one could do without the coupons with value -1, because it is possible for either player to pick an extremely small value of δ. However, the -1 point coupons make it easier to sell the concept of coupon stacks to serious competitive gamesmen. A good environment to accompany a 10×10 game of Amazons needs almost 80 coupons with value -1, even though δ of 0.1 or even 0.5 proves interesting. If there were few or no -1 point coupons, the appropriate δ might need to be reduced by nearly two orders of magnitude.

take the top coupon from the stack. Even if the value of the current position, G, becomes a number, play continues until after only -1 point coupons remain. Play terminates after each of the players has taken three -1 point coupons consecutively. Then there is a concluding ceremony, during which the referee gives a special $-\frac{1}{2}$ point *terminal komi* coupon to the player who did not take the last of the six consecutive -1 point coupons which caused the game to be terminated. Each player's score is then computed as the sum of all of the coupons he has taken. If the magnitude of the difference between these scores exceeds ε, the player with the higher score is declared the winner. Otherwise, the result is declared to be a tie.

Comment. *Why do we not end the game until after the sixth consecutive coupon of value -1 is taken? In part, this is modeled after a well-known rule in chess, which declares the game to be terminated with a drawn outcome after the same position occurs three times with the same player to move. Presumably this is intended to give each player multiple chances to consider other options. Theoretically, if both players are playing optimally, one might think that the first repetition of a position would be sufficient. However, in games like Go, which include a ko rule, there are situations in which a pair of consecutive coupons is taken as a kothreat and its response while the game is still quite active. So we might theoretically weaken the six consecutive coupons to four consecutive coupons but, at least in the case when the board positions include Go, two consecutive coupons is definitely not enough.*

3. Definitions of Idempotents Without Opening Ceremonies

On. Figure 1 shows two positions of a Black checker king. Although White has no pieces to move, Black can move to and fro between these two positions whenever he so desires. I call these positions *onto* and *onfro*. From *onto*, Black can move to *onfro*. From *onfro*, he can move back to *onto*. Formally,

$$\mathbf{onto} = \{\mathbf{onfro} \mid \}, \qquad \mathbf{onfro} = \{\mathbf{onto} \mid \}.$$

Taken together, **onto** and **onfro** can be viewed as the bifurcated components of an abstract game called **on**, which appears in the latter half of Chapter 11 of *Winning Ways*. In a sum of games including an **on**, Right will never be able to

Figure 1. Onto (left) and onfro (right).

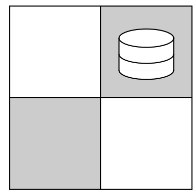

Figure 2. Off.

force the play of the game to terminate. Formally,

$$\mathbf{on} = \{\mathbf{on} \mid \}.$$

This game has infinite mean. If X is any traditional game, then our definition of equality implies that

$$\mathbf{on} + X = \mathbf{on}.$$

Off. **Off** differs from **on** only in that the lone checker king is now white instead of black. Now it is Right who can move to and fro at will. In a sum of games including an **off**, Left will never be able to prevent Right from playing. If Left has only a finite number of moves available elsewhere, he will eventually run out of options and lose the game.

This game has mean value of $-\infty$. Like **on**, it overpowers any traditional game to which it is added.

Dud. Although **on** and **off** superficially look like negatives of each other, every checker player knows that their sum is not zero (a win for the second player), but a *Deathless Universal Draw*, often realized on a single checkerboard when each player has only a single king, and the two kings are located in the double corners at opposite ends of the board.

Dud is not only a draw by itself, it also ensures that anything to which it is added also becomes a drawn game. **Dud** plays the role of a black hole. If a dud is present, no other features matter; the overall game behaves as a **dud**.

Comments on Outcomes with no Winner. Terminology has some history. In *Winning Ways*, we distinguished between *ties*, in which play terminated without either player winning, and *draws*, in which play could naturally be *drawn* out forever if not terminated by a special rule. In chess tournaments, ties due to stalemate and draws due to perpetual check are treated identically: each player receives $\frac{1}{2}$ win + $\frac{1}{2}$ loss. However, Go has a quite different tradition. The natural rules allow certain positions to be repeated immediately in a two-move loop, one move by each player. Such global repetitions are universally banned by the so-called *Ko rule*. Positions involving ko are very common, occurring several times in most games. Lengthier repetitions, after 4 or 6 or more moves are also possible, but rather rare, occurring in only a small fraction of one percent of all professional games. The rules about how to handle such *superko* positions differ from time to time and from place to place. Today, only the North American rules and the New Zealand rules simply ban all superkos. Japanese rules explicitly allow them. Chinese and Taiwanese rules for superkos are more complicated. In many cases, one player or the other is permitted to repeat the position but the other player is not. Which player is banned depends on the details of the position.

If a game gets hung in a superko, the Japanese tournament rules do NOT treat it as a tied or drawn outcome. The official translation defines the outcome as *no result*. So I prefer the English word *hung*, as in a hung jury or a computer program that is hung in a loop. Like the hung jury, a Japanese Go game which hangs in a superko often leads to a new game in which the same two contestants begin again from scratch.

To further complicate the situation, some amateur Japanese Go games can end with a tied score. Many translators have called such an outcome a *draw*, in direct conflict with the terminology of *Winning Ways*. I call them ties, and try to avoid any use of the word *draw* in reference to Go. As we are primarily concerned with individual games, or sums of games, the question of how such outcomes are treated in tournaments need not concern us here. So chess games which draw in perpetual check and checker games which draw in a **dud** might also be said to have *hung*. Whatever the terminology, a win is surely better than either a tied or hung outcome, which in turn is better than a loss.

Loony. Some impartial games have what are called *complimenting moves*, and such games can have positions with a fascinating value called *loony*, and denoted by the lunar symbol, ☽. One of the best-known such games is Dots-and-Boxes, whose Impartial variation I shall now describe.

The game is played on a array of dots located on the integer points of a rectangular subset of the Cartesan plane. These dots appear at unit distances from each other in vertical and horizontal rows. A legal move for either player is to draw a new horizontal or vertical line of length one, joining two dots. Unless that line completes a unit square, it completes the mover's turn. However, if that line completes one or two unit squares (called boxes), the mover must continue to make another move. The game ends when no further legal moves remain, and the player who made the last move *loses*.

(Even though last player loses, this is regarded as a *normal* rather than a *misère* rule because the last move necessarily completes a box, so the turn is incomplete. The player loses because he is unable to fullfill the requirement that he make another move.)

Figure 3, left, shows the position of an impartial Dots-and-Boxes position. This position can be viewed as the sum of four positions: two squares in the northwest, one in the northeast, two in the southeast, and four in the southwest, whose respective values can be shown to be ☽, $*$, $*$, and $*2$. The figures in the middle and on the right show two quite different ways in which the player to

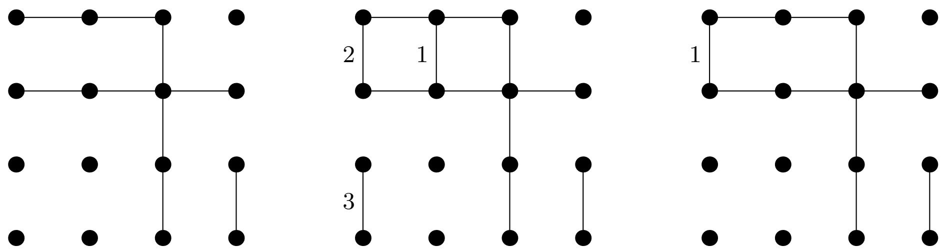

Figure 3. A Dots-and-Boxes position (left) and two possible continuations.

move can complete his turn. In the middle, he takes each of the boxes in the northwest and then completes his turn with a move elsewhere. In the right-hand figure, he completes his turn in another way, which forces his opponent to make the first move outside the northwest region. So although it might not be easy to determine whether or not one wishes to play first or second on the rest of the board, it is easily seen that the player to move from a sum which includes a loony component can win the game in either case. If he wants to make the first move elsewhere, he plays the loony component in a way which enables him to do that, as in the middle figure. On the other hand, if he wishes to force his opponent to make the first move elsewhere, he can achieve that objective by playing the loony component in the other way.

Loony values can also occur in games with entailing moves, as described in *Winning Ways*, Chapter 12. Entailing moves are special moves that require the opponent to move in a certain portion of the game. Complimenting moves are special moves (like completing a box in Dots-and-Boxes) which can be viewed as forcing the opponent to skip his next turn. Rather than attempt any general definition, for purposes of this paper it is sufficient to simply define ☽ very specifically as the northwest corner of the impartial Dots-and-Boxes position of the left panel in Figure 3.

4. The Addition Table

The addition table for the idempotents we have discussed is shown on the next page. The order in which they are listed may be viewed as an order of increasing *vim* (as in "vim and vigor"), with the understandings that **on** and **off** have equal vim, and that otherwise, whenever two of the idempotents are added, the one with the more vim predominates.

Another view is that adding in any of these idempotents destroys certain information, and the idempotents with more vim are more destructive.

In practice, adding in an appropriate idempotent can often be the key to the analysis of a particularly challenging position. The most helpful idempotent is the one which preserves just those features which are crucial to the winning line of play, while annihilating all of the less significant considerations.

Addition of Idempotents

zero	0	☽	☆	ish	$\mathcal{E}_\tau$	$\mathcal{E}_t$ $(t>\tau)$	$\mathscr{E}_t$	**on**	**off**	**dud**
loony	☽	☽	☆	ish	$\mathcal{E}_\tau$	$\mathcal{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
remote star	☆	☆	☆	ish	$\mathcal{E}_\tau$	$\mathcal{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
	ish	ish	ish	ish	$\mathcal{E}_\tau$	$\mathcal{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
enriched environments	$\mathcal{E}_\tau$	$\mathcal{E}_\tau$	$\mathcal{E}_\tau$	$\mathcal{E}_\tau$	$\mathcal{E}_\tau$	$\mathcal{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
enriched environments	$\mathcal{E}_t$ $(t>\tau)$	$\mathcal{E}_t$	$\mathcal{E}_t$	$\mathcal{E}_t$	$\mathcal{E}_t$	$\mathcal{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
enriched environments	$\mathscr{E}_t$	$\mathscr{E}_t$	$\mathscr{E}_t$	$\mathscr{E}_t$	$\mathscr{E}_t$	$\mathscr{E}_t$	$\mathscr{E}_t$	**on**	**off**	**dud**
	on	**on**	**on**	**on**	**on**	**on**	**on**	**on**	**dud**	**dud**
	off	**off**	**off**	**off**	**off**	**off**	**off**	**dud**	**off**	**dud**
	dud	**dud**	**dud**	**dud**	**dud**	**dud**	**dud**	**dud**	**dud**	**dud**

5. Other Properties of These Idempotents

Homomorphisms. Each idempotent I induces a mapping: $A \mapsto A+I$. If also $B \mapsto B+I$, it is evident that $A+B \mapsto A+B+I$. So each such mapping is a homomorphism on the semigroup $\mathcal{S}$.

We now give expanded descriptions of some of the idempotents:

Loony. Loony, or ☽, lies at the heart of a rich structure described in some detail in [Berlekamp 2000b]. This context, within which ☽ seems so powerful, is impartial; the only other games which appear there are nimbers. When compared with partesan games, we soon discover that ☽ has so little vim that it preserves *all* strict inequalities in the ordering relationships within the traditional group, $\mathcal{G}$. To see this, it is sufficient to verify that for every positive integer n,

$$\mathbf{miny}_n < \text{☽} < \mathbf{tiny}_n$$

where $\mathbf{tiny}_n$ is defined as $0 \,\|\, 0 \mid -n$ and $\mathbf{miny}_n = -\mathbf{tiny}_n$. It has long been known (as can be verified by an induction on zero-based canonical birthdays) that every positive game with a finite number of positions is greater than tinies which have sufficiently large n. So, if G and H are traditional games for which $G > H$, the inequality is preserved if ☽ is added to either or both sides.

Remote Star. Adding $\star$, a remote star, into a game induces a homomorphism whose kernel contains all infinitesimals of higher order than $\uparrow = 0 \mid *$.

Remote star plays the central role throughout Chapter 8 of *Winning Ways*, which presents the theory of atomic weights, including uppitiness and the Norton multiply operation. When the results of that chapter were first discovered, the main application was the game of Hackenbush hotchpotch, where the remote star, and its proximity to the red and blue kites, arises quite naturally. The generalization to other "all-small" games such as childish Hackenbush was fairly immediate, although we neglected to notice the relevance to the Toads-and-Frogs analysis in an earlier chapter of *Winning Ways*. Another decade passed before the discovery of the crucial role which atomic weight theory plays in getting the last one-point move in Go endgames [Berlekamp and Wolfe 1994; Moews 1996]. One must wonder why atomic weight theory, or even some approximate version of it, eluded Go players for so many centuries. I think the explanation is that, except for $*$, the nimbers (of which the remote star might be viewed as a limit) occur only very rarely in Go or in chilled Go.[2] So here we have a good example of how an issue in one game (How do I get the last one-point move in Go?) benefits from insights that originate in another quite different game (Hackenbush hotchpotch).

Enriched Environments. We shall see that adding a sufficiently hot enriched environment preserves the mean, but destroys all other information about the traditional game to which it is added. This is why it has become such an important tool in the study of Amazons [Berlekamp 2000a, Snatzke 2002] and Go: It eliminates all of the less essential features, allowing one to focus on the primary attribute of each position. At a relatively early stage of the endgame, it enables one to obtain a real-valued *count* which is precisely accurate in a well-defined mathematical sense. In many cases, the board contains numerous small battles which themselves serve as a plausible approximation to the enriched environment. For this reason, the line of play which is provably optimum in the enriched environment often provides an extraordinarily good guide as to how to play the actual endgame.

Coupons with positive values behave exactly like switches. The only real difference between $\mathcal{E}$ and $\mathscr{E}$ is that the former stack of coupons ends at 0, whereas the latter stack also includes negative coupons and a terminal komi. In other words, for positive t,

$$\mathscr{E}_t = \mathcal{E}_t + \mathscr{E}_0.$$

When $\mathcal{E}$ is played all by itself, the players alternately take coupons, effectively dealing the stack out between them. Each player ends with $n/2$ coupons, and each of the second player's coupons is δ less than the prior coupon just taken

[2]Even the existence of $*2$ was until very recently an open problem (listed for example in [Guy 1996, p. 487, problem 45]); its resolution came in [Nakamura and Berlekamp 2002] in the form of a chilled Go position of value $*2$.

by his opponent. So first player wins by a net score of $n\delta/2 = t/2$. One way to make the game fairer is to follow the common practice used in professional Go games, which assigns a specified number of points to the second player (White) as compensation for the advantage which Black gets from the first move. This compensation is called the *initial komi*. Evidently, the fair initial komi corresponding to $\mathcal{E}_t$ is $t/2$. Similarly, it can be verified that the terminal komi (of value $-\frac{1}{2}$) included in the coupon stack $\mathscr{E}_0$ is just enough to balance out the disadvantage of the last coupon (at temperature -1), so that when $\mathscr{E}_0$ is played out all by itself, the net score has magnitude less than ε, so the resulting outcome is a tie.

Thermographs. The thermograph of a given traditional game, G, consists of two functions of an independent variable, t, which is called the temperature. The functions are the thermograph's Leftwall and Rightwall, which may be defined as follows:

$$\text{Leftwall}(G,t) = \text{Lscore}(G + \mathcal{E}_t,\ t) - t/2,$$

$$\text{Rightwall}(G,t) = \text{Rscore}(G + \mathcal{E}_t,\ t) + t/2.$$

It can be shown that for traditional games, these definitions yield the same walls of the thermograph as the recursions based on the cooling homomorphism found in such references as Chapter 6 of *Winning Ways*. Alternatively, if one takes those recursions as the definition, then one can prove that the scores of $G+\mathcal{E}_t$ are the solutions of these same equations. In other words, the final score of a well-played game consisting of a traditional game, G, plus an enriched environment is equal to the fair initial komi, $t/2$, plus whichever wall of the thermograph corresponds to whoever plays first.

It is conventional to plot thermographs with the independent variable, t, running upwards along the vertical axis, starting from a minimal value which is often taken to be 0 or -1. Positive values are plotted as increasing towards the left along the horizontal axis, negative values to the right. This reversal of the signs normally used in analytic geometry facilitates a more direct comparison between thermographs and the expressions or graphs of the underlying games, in which Left seeks to move leftward, and Right seeks to move rightward.

If t is sufficiently large, then $\text{Leftwall}(G,t) = \text{Rightwall}(G,t) = \text{Mast}(G)$, a value independent of t. Traditionally, this mast is also known as the *mean value*, or mean(G). The greatest lower bound on temperatures, τ, such that

$$\text{Leftwall}(G,t) = \text{Rightwall}(G,t) = \text{Mast}(G) \text{ for all } t \geq \tau,$$

is called the *temperature* of G.

In some applications, it is useful to color portions of the mast. If $t > \tau$ and if Left can attain the optimal Leftscore of $G + \mathcal{E}_t$ by playing on G, the mast of G at t can be colored *blue*. Similarly, if Right can attain the optimal Rightscore of $G + \mathcal{E}_t$ by playing on G, the mast of G at t can be colored *red*. A portion of the mast that is colored both red and blue is shown as *purple*. A portion that is neither red nor blue is shown as yellow (or gray). All traditional games have thermographs whose masts become gray at all sufficiently high temperatures.

One can also apply this reasoning to a sum of traditional games, $A + B + C$. For large t, one can study $A+\mathcal{E}_t$, $B+\mathcal{E}_t$, $C+\mathcal{E}_t$, and then add them up to obtain

$$\begin{array}{r} A + \mathcal{E}_t \\ B + \mathcal{E}_t \\ C + \mathcal{E}_t \\ \hline A + B + C + \mathcal{E}_t, \end{array}$$

because $\mathcal{E}_t$ is an idempotent. Evidently, $\mathrm{Mast}(A+B+C) = \mathrm{Mast}(A)+\mathrm{Mast}(B)+\mathrm{Mast}(C)$, a result well-known in traditional thermography.

If G is any traditional game and n is any positive integer, then $\mathrm{Mast}(n\,G) = n\,\mathrm{Mast}(G)$ and $\mathrm{temp}(nG) \geq \mathrm{temp}(G)$. The scores of traditional games can be bounded by

$$-\tau + \mathrm{Mast}(G) \leq \mathrm{Rightscore}(G) \leq \mathrm{Mast}(G),$$

$$\mathrm{Mast}(G) \leq \mathrm{Leftscore}(G) \leq \tau + \mathrm{Mast}(G),$$

$$\mathrm{Mast}(G) - \frac{\tau}{n} \leq \frac{\mathrm{Scores}(nG)}{n} \leq \mathrm{Mast}(G) + \frac{\tau}{n},$$

from which is becomes reasonable to say that $\mathrm{Mast}(G)$ is the *mean* of G.

Thermographs of Games Including Kos. A common feature of Go positions is a situation called *ko*, which is a 2-move loop in the game graph of a local position. Globally, the immediate repetition of a position is banned in all dialects of the rules of Go. However, since the ban is global, it is quite common for the local position to be repeated locally in a so-called *kofight*. During the kofight, many moves elsewhere become worthwhile as *kothreats*. Locally, the game can stay in a two-cycle loop as long as each player is able to find worthwhile threats elsewhere.

The recursive definitions which formed the original basis of thermography were not designed to handle the loopiness of kos. This problem was addressed by Berlekamp [1996]. That paper introduced an extended thermography which was subsequently applied to the study of a collection of over eighty kos by Berlekamp, Müller and Spight [Berlekamp et al. 1996]. Many of those kos were taken from professional games studied by Müller.

Even though the play of actual kos can depend significantly on kothreats located in other regions of the board, the total environment consisting of all of these kothreats can often be usefully approximated by one of a small number of possibilities, depending on who is *komaster* or *komonster*, with a relatively small intermediate *hypersensitive* region. When the mast of a game depends on who is komaster of a ko occurring in one the game's positions, that game is said to be *hyperactive*. Other positions may contain kos which are called *placid*, because their means are independent of who is the master of the ko.

In any sum of games including at most one hyperactive ko position in which it is clear who is komaster, it remains true that the mast of the sum is still the sum of the masts. Because of this, extended thermographs prove very useful for analyzing a wide range of positions from actual Go games, which typically have at most one position depending on a hyperactive ko. However, sums of two or more games dependent on hyperactive kos are more complicated.

If a position's mast depends on a hyperactive ko, then its *mean* typically differs from either its mast with Left komaster or its mast with Right komaster. In many cases, the analysis of the game depends on one of these two masts rather than on the mean.

Spight [1998] has begun a theoretical investigation of sums of several hyperactive positions. The situation is complicated. Fortunately such sums are not very common in practice.

In some situations, an analysis of the whole board may be needed to distinguish whether a game including a ko position is best characterized as Left komonster, Left komaster, hypersensitive, Right komaster, or Right komonster. But often the answer is heavily biased by factors which depend only on a careful local analysis of the hyperactive ko.

Studies of this topic are continuing to progress at a rapid rate, as discussed in the section on further work at the conclusion of this paper.

Positive or Full Enrichment? Amazons [Berlekamp 2000a] is a hot game which, like Go, is primarily a territorial battle. However, unlike Go, positions with the traditional *number* values do occur in Amazons; Snatzke [2001] recently discovered a value of $\frac{1}{16}$. But even the best current players are often unable to determine whether the value of a position is a number or not. So enforcement of the traditional *stopping* rule of *Winning Ways*, pp. 145–162 is problematical. Negative coupons were introduced to address this issue.

The negative coupons in $\mathscr{E}$ have also proved helpful in teaching subzero thermographs [Berlekamp 1996]. Game positions with subzero thermographs actually occur in chilled Go [Berlekamp and Wolfe 1994], and some of their properties are more easily understood when chilled rather than when warmed [Takizawa 2001].

Environmental Go The concept of enriched environments grew out of a sequence of attempts to improve communication with professional Go players.

Historically, thermographs were first developed in terms of taxes rather than stacks of coupons. However, efforts to interest serious Go players in formal definitions of cooling and heating were very unsuccessful. The notion of determining tax rates by a competitive auction in [Berlekamp 1996] failed to catch on. Most people simply don't like to think about taxes. Changing the sign of the payoff (at positive temperatures) from negative to positive, and calling these payoffs *coupons* instead of *taxes* had a large impact. Jujo Jiang and NaiWei Rui, both famous 9-dan Go players, agreed to play the first demonstration game of environmental Go in 1998. An analysis of the endgame of this game appears in [Spight 2001]. Jiang and Rui played another game of environmental Go at MSRI on July 23, 2000. The analysis of that game is still underway. Many participants at the American Go Congress held in Denver in August 2000 expressed much interest in Environmental Go. As a means to provoke quantitative discussion, the concept of a stack of coupons must now be viewed as an enormous success.

The most popular current stack consists of 40 coupons, all positive, with values 20, 19.5, 19, ..., 0.5 This stack is placed next to the initially empty Go board. White receives a komi of 9.5 points, and Black moves first. There is universal agreement that Black should open by taking the 20-point coupon. White responds by taking the 19.5-point coupon. Excitement mounts as the play continues. Eventually, when the temperature of the coupon stack descends to somewhere in the low teens, someone plays the first stone on the board. Possibly the opponent replies by taking another coupon. Outstanding professionals such as Jiang and Rui often provide fascinating games. Unlike a conventional Go game, the environmental game forces the players to give us their expert opinions, at every move, as to whether or not the current move on the board is worth as much as the top coupon on the stack. In this way, we extract some interesting quantitative expert opinions about how big the various moves are, at least to within the $\delta = 0.5$ point difference between successive coupons.

Of course, the popular coupon stack is only a crude approximation to $\mathscr{E}_{20}$. One approximation is that $\delta = 0.5$ is relatively large. Another is that the popular initial komi is 9.5, which differs from the fair komi of $\mathscr{E}_{20}$, which is 10. These discrepencies are the results of an effort to maintain as much compatibility as feasible with the traditions and practices of conventional Go, where the score excluding the komi is necessarily an integer, and where the komi is a half-integer (to avoid ties) and generally one more than an *odd* integer (in order to minimize the risk that Chinese and Japanese scoring systems might disagree on the outcome of the game).

In Go, it is very difficult to realize a traditional game value of $\frac{1}{2}$. Bill Spight [2000] has constructed a position including a hyperactive ko and two sekis which, if there are no other kothreats on the board, behaves like the traditional mathematical value of $\frac{1}{2}$. Possibly the game value of $\frac{1}{2}$ cannot be realized on a Go board without a hyperactive ko position. I know no proof of this, but if such a position were to occur, it would almost certainly result in a dispute. After an appeal to the rules committees, this position would find its way into the collections of exotic cases covered by special scoring rules. Such collections now appear as appendices to the various dialects of official Go rules used in various Asian jurisdictions [Bozulich 1992]. The fact that none of the positions found in these appendices has traditional game value of $\frac{1}{2}$ provides strong evidence that such values are either impossible or can be constructed only with great effort.

Some of the more common complexities which can occur in Go scoring are described in Chapter 8 of [Mathews 1999], an excellent introductory book. Exotic subtleties substantially more intricate than have ever appeared in any known professional game have been composed by Harry Fearnsley [2000a; 2000b].

Most professional Go players are quite adept at calculating means and temperatures of latestage endgame positions. Their methodology differs somewhat from ours. It is faster but less accurate. Under appropriate assumptions which are usually satisfied in practice, it yields answers which agree with ours when the position is sufficiently simple and no kos are relevant. When the position is more complicated, professionals usually get approximate values quickly. Compared with mathematicians, the Go pros are vastly better at seeing the best local moves, but less patient and persistent in comparing the temperatures of different moves in different regions. Pros do approximations very quickly, but the mathematicians achieve more quantitative accuracy, and in many cases this improved accuracy translates back to improved lines of play which the players readily appreciate.

Orthodox Moves. Suppose that an enriched environment of sufficiently high temperature, t, is added to a game G, minus its mast value, and plus the appropriately signed fair initial komi, $t/2$. Then if both players play optimally, the net score at the end of the game will have magnitude less than ε, and the game will be declared a tie. The moves which a guru might play in the course of such a game are called *orthodox*. Other moves are called *unorthodox*.

As illustrated in [Berlekamp 2000a], there are many situations in which the traditional *canonical* methodology leads to a creeping growth of unmanageable complexity. So rather than consider all options of G which might be useful in playing $G + X$ for *some* X, orthodoxy focuses on playing $G + X$ for one particularly tractable value of X, namely $X = \mathcal{E}_t$. Thanks to the fact that $\mathcal{E}$ is idempotent, orthodox moves in G and H will, at the appropriate temperatures, remain orthodox moves in $G + H$.

An algorithm called *sentestrat* provides advice on how to play a sum of games. When a stack of coupons is present, the *ambient temperature* is defined as the value of the top coupon. If the opponent has just moved to a board position whose temperature is now higher than the ambient, then sentestrat tells you to respond in the same local game wherein your opponent has just moved. Otherwise, play in whichever region of the board or stack has the highest local temperature. Sentestrat is provably an orthodox strategy, assuming the only hyperactive ko (if any) has a clear komaster.

The *board temperature* is the temperature of the hottest region in which it is legal to move. (When a koban is applicable, the temperature of the ko is excluded.) The board temperature of a sum is the maximum of the temperatures of the summands.

One might choose to play an orthodox strategy, behaving as if coupons were present, even when there is no stack of coupons available. This requires a revised definition of the ambient temperature, which is defined historically. It is the minimum board temperature of any position which has yet occurred.

In the presence of an enriched environment, orthodox play by either player ensures that he will attain a result as least as desirable as the orthodox score, which is the mast value adjusted by the fair komi. When no enriched environment is present, then there are situations in which an unorthodox strategy can do better. Finding the optimal lines of play in such situations may require considerable search. The next section discusses an accounting technique for evaluating candidate lines of unorthodox play.

Orthodox Accounting. Orthodox accounting quantifies the benefits and costs of each move, whether the move is orthodox or not. The system is based on the forecasting methodology presented and illustrated in [Berlekamp and Wolfe 1994]. The present version includes refinements which handle kos.

Orthodox accounting attributes the final net score of a game to five different types of terms:

1. Mast value of the current position.
2. Fair current komi. The magnitude of this term is one half the ambient temperature, and the sign favors whoever's turn it is to move next.
3. $\frac{1}{2}$ Summation of signed temperature drops. Whenever the ambient temperature drops by Δt, then half that drop is awarded to whichever player got the last move at the old (higher) temperature.
4. Komaster adjustments. In many cases, these have already been correctly accounted for in term 1.
5. Komonster adjustments. This occurs only when one player finishes play on a hot ko; he then gains an adjustment equal to the difference between the temperature of his move and the ambient temperature.

In some sense, orthodox accounting is like a bank statement; it allows the observer, after the fact, to see exactly what credits came in and what debits were paid. It is a good accounting system in the sense that it always starts and ends with the correct balances, and the current *bottom line* does not undergo spurious big swings. In particular, suppose an orthodox move changes the mean by the current ambient temperature, which remains unchanged. In such a *transaction*, the changes in terms 1 and 2 cancel out, and the predicted net score remains unchanged.

When an enriched environment is present, Terms 3 and 5 are negligible. To see this, let m be an upper bound on the total number of moves that might be played on the board. For example, in a 19×19 Go game, most players would regard $m = 400$ as such a bound. Then since either player can place a bid which ensures that $\delta < \varepsilon/2m$, we can assume that at least one of them does so. In the course of the game, term 3 will provide many adjustments, but the number favoring one side cannot differ from the number favoring the other by more than m. A similar argument can bound the total effect of terms 5. When one player becomes komonster, he plays a kothreat and gets a response before retaking the ko. His opponent then takes a coupon as a kothreat, to which the komonster responds by taking another coupon after which the komonster's opponent retakes the ko. So the net affect of this typical sequence of six moves is two moves on the board and a decrease in ambient temperature of 2δ. So the sum of the magnitudes of all terms of type 5 cannot exceed $m\delta$.

The presence of an enriched environment ensures that the ambient temperature will decrease adiabatically. When no enriched environment is present, a decrease in the ambient temperature can be larger than δ. Such a decrease is called a *thermal shock*. Although an individual thermal shock can be significant, the sum of the magnitudes of all terms 3 is precisely the same as the magnitude of the fair initial komi, which is $t/2$. It is possible for the same player to get the benefits of all thermal shocks. But it is much more common for these benefits to be nearly evenly divided between the two players.

No matter how impoverished the environment, nor how poorly the game is played, orthodox accounting assigns a precise cost to each unorthodox move. In order for such a move to be a wise investment, it must lead to a future payoff via an item of type 3, 4, or 5. Even when an enriched environment is present, it is possible for the costs of the unorthodox move to be justified if it changes the balance of kothreats in such a way as to change the master of a forthcoming hyperactive ko. In an impoverished environment, there can also be other opportunities for returns on unorthodox investments (see the section on suspense and remoteness in Chapter 6 of *Winning Ways*). However, in the professional Go games we have analyzed, many of the unorthodox moves we have identified have simply turned out to be mistakes. Others subsequently breakeven in the sense that the cost is later recovered and the final score is the same as it would have been if an orthodox line had been played. Cases where unorthodox moves actually yield a profit seem to occur quite rarely in real play.

6. Suggestions for Further Work

Analysis of Real Go Endgames. The combination of local orthodox analysis, plus orthodox accounting, plus well-known search techniques such as alpha-beta pruning (which are well known in the artificial intelligence community) looks very promising but has yet to be thoroughly investigated and implemented.

Refinements of Ish. One can envision an expansion of the idempotent addition table to include more elements. Just below loony might be some sort of generic **tiny**. Tiny is a symbol which becomes completely defined only when followed by a subscript, G, which must be a game that is larger then some positive number.

Then $\mathbf{tiny}_G = 0 \,\|\, 0 \mid -G$. The smallest traditional tinies with finitely many positions are those for which $G = n$, a very large integer, and the largest are those for which $G = 2^{-k}$, a very small positive fraction. In some games we frequently encounter tinies whose subscripts are very complicated games which are definitely not numbers. Several of the positions in the complete analysis of Toads-and-Frogs in *Winning Ways*, Figure 12, p. 132 of are of this type. Chilling almost any professional Go game yields several such positions. So, it is often convenient to treat *tiny* generically, with the subscript unspecified. One may then need to specify whether the assertion one is making applies to all possible values of the subscript or to only some possible values of the subscript. In many cases, it doesn't matter, as in the assertion earlier in this paper that

$$\mathbf{miny} < ☽ < \mathbf{tiny}$$

So I have yearned to insert **tiny** into the addition Table between ☽ and ☆, but I have not found any plausible opening ceremony with which to define a generic tiny that would be as nice as the other idempotents, and compatible with them.

In chilled Go, it is often desirable to distinguish between

$$0 \mid \mathbf{tiny} + \downarrow$$

and

$$0 \mid \mathbf{tiny} + \downarrow *$$

Both are common. The latter is a positive infinitesimal much bigger than tiny, but still small enough to be negligible for most purposes. The former is closely approximated by $*$. One may need to preserve the distinction between $*$ and 0 while neglecting $0|\mathbf{tiny} + \downarrow *$. The remote star is too crude for this purpose, because

$$☆ + *n = ☆$$

for all n, including $n = 1$.

An infinitesimal which preserves distinctions between 0 and $*$ needs not only to be of higher order than $\uparrow$, but also of higher order than $\uparrow^n$ [Conway 1976, pp. 199–200].

Near the big end of the range of ish, one might try to retrieve at least the biggest term in Norton's thermal dissociation (described in the Heating section of *Winning Ways*, Chapter 6). This term is nicely preserved by traditional *cooling*, but eradicated by enriched environments. In some composed problems, these terms can point the way to low-cost unorthodox plays which capture the benefit of a big thermal shock.

Kothreat Environments. In August 2000, Spight, Fraser, and I began studying several promising models of *kothreat environments*. These environments all behave like 0 when added to any traditional loopfree game, but they can have desirable simplifying effects on kos.

References

[Berlekamp 1996] E. Berlekamp, "The economist's view of combinatorial games", pp. 365–405 in *Games of no chance* (Berkeley, 1994), edited by R. Nowakowski, Math. Sci. Res. Publ. **29**, Cambridge Univ. Press, Cambridge, 1996.

[Berlekamp 2000a] E. Berlekamp, "Sums of $2 \times N$ Amazons", pp. 1–34 in *Game theory, optimal stopping, probability and statistics*, edited by T. Bruss and . LeCam, Lecture Notes – Monograph Series **35**, Inst. of Math. Stat., Hayward, CA, 2000.

[Berlekamp 2000b] E. Berlekamp, *The-dots-and boxes games: sophisticated child's play*, A K Peters, Natick (MA), 2000.

[Berlekamp 2002] E. Berlekamp, "The 4G4G4G4G4 problem and its solution", in *More games of no chance*, edited by R. Nowakowski, Math. Sci. Res. Publ. **42**, Cambridge Univ. Press, Cambridge, 2002.

[Berlekamp and Wolfe 1994] E. Berlekamp and D. Wolfe, *Mathematical Go: Chilling gets the last point*, A K Peters, Wellesley (MA), 1994.

[Berlekamp et al. 1982] E. R. Berlekamp, J. H. Conway, and R. K. Guy, *Winning ways for your mathematical plays*, Academic Press, London, 1982. Second edition, A K Peters, Natick (MA), 2001.

[Berlekamp et al. 1996] E. Berlekamp, M. Müller, and W. Spight, "Extended thermography for multiple kos in Go", Technical Report 96-030, Berkeley, 1996.

[Bozulich 1992] R. Bozulich, *The Go player's almanac*, Ishi Press, Mountain View, CA, 1992.

[Conway 1976] J. H. Conway, *On numbers and games*, London Math. Soc. Monographs **6**, Academic Press, London, 1976. Reprinted by A K Peters, Natick (MA), 2001.

[Fearnsley 2000a] H. Fearnsley, 2000. See http://www.goban.demon.co.uk/go/main.html.

[Fearnsley 2000b] H. Fearnsley, 2000. See http://www.goban.demon.co.uk/go/bestiary/zippersetc.html.

[Guy 1996] R. K. Guy, "Unsolved problems in combinatorial games", pp. 475–491 in *Games of no chance* (Berkeley, 1994), edited by R. Nowakowski, Math. Sci. Res. Publ. **29**, Cambridge Univ. Press, Cambridge, 1996.

[Mathews 1999] C. Mathews, *Teach yourself Go*, Hodder and Stoughton, London, 1999.

[Moews 1996] D. Moews, "Infinitesimals and coin-sliding", pp. 135–150 in *Games of no chance* (Berkeley, 1994), edited by R. Nowakowski, Math. Sci. Res. Publ. **29**, Cambridge Univ. Press, Cambridge, 1996.

[Nakamura and Berlekamp 2002] T. Nakamura and E. R. Berlekamp, "Analysis of composite corridors", ICSI technical report, 2002. See http://www.icsi.berkeley.edu/techreports/2002.abstracts/tr-02-005.html.

[Snatzke 2001] R. G. Snatzke, "Exhaustive search In Amazons", in *More games of no chance*, edited by R. Nowakowski, Math. Sci. Res. Publ. **42**, Cambridge Univ. Press, Cambridge, 2002.

[Spight 1998] W. Spight, "Extended thermography for multiple kos in Go", pp. 232–251 in *Computers and Games* (Tsukuba, Japan, 1998), edited by H. J. van den Herik and H. Iida, Lecture Notes in Comp. Sci. **1558**, Springer, Berlin, 1998.

[Spight 2000] W. Spight, 2000. Unpublished correspondence.

[Spight 2001] W. Spight, "Go Thermography: The 4/21/98 Jiang-Rui Endgame", in *More games of no chance*, edited by R. Nowakowski, Math. Sci. Res. Publ. **42**, Cambridge Univ. Press, Cambridge, 2002.

[Takizawa 2002] T. Takizawa, "An application of mathematical game theory to go endgames: some width-two-entrance rooms with and without kos", in *More games of no chance*, edited by R. Nowakowski, Math. Sci. Res. Publ. **42**, Cambridge Univ. Press, Cambridge, 2002.

Elwyn Berlekamp
Mathematics Department
University of California
Berkeley, CA 94720-3840
berlek@math.berkeley.edu