COMPUTER JOTTO AND CRASH

LEONARD GORDON
Tucson, Arizona

The August 1996 article "Playing Jotto Against a Computer" discusses the game of Jotto as played by a commercial computer program. Toward the end of the article, the authors suggest a "better" (minimax) value of play. I agree that their method is too difficult to implement, but I also believe that it is based on a weak premise and would be useless even if implemented. There can be no compleat strategy to Jotto. Guessing your opponent's word is a matter of bookkeeping plus a little luck. Since the computer is a better bookkeeper and knows its vocabulary better than a human opponent, its best strategy is thoughtful random play.

To test my opinions, I wrote a computer program. (As always happens, I got carried away and wrote several.) My programs do not play a full game; they only guess an opponent's target word. While testing my programs. I noticed certain weaknesses in the logic (places where a clever opponent could force the computer into an excessively long search) and introduced a few remedies; after that the programs seemed to play pretty well. (How well? The reader must decide from reading this article, or from trying out my programs himself, as offered in the November 1996 Colloquy.) After Jotto, I wrote a program for a form of Giotto (14th-century Italian painter) best known to Word Ways readers as Crash. Both programs use five-letter Official Scrabble Players Dictionary words.

THE JOTTO ALGORITHM

My algorithm accomplishes the same thing as the one described in the August article. Establish a linear array, V(L), for L = 1 to 26. Add to it after every word guessed. For example, if the first guess is CRWTH, we mark V(3) = 1, V(18) = 1, V(23) = 1, V(20) = 1, V(8) = 1, Say the second word is ORGAN. We now add V(15) = 6, V(18) = 1 + 6 = 7, V(7) = 6, V(1) = 6, V(14) = 6. If our third guess is STARE, we add V(19) = 36, V(20) = 1 + 36 = 37, etc. At the 12th guess we add 362,797,056 (6 to the eleventh power) per letter.

From reports of the number of hits, we compile HSUM = H1 + 6H2 + 36H3 + 216H4 + ..., where H(i) is the number of hits scored by the ith guess word. The first few guesses by the computer may be preprogrammed (independent of the hit numbers); however we eventually get to the point where we want to calculate the value of each potential next guess word using VAL = V(L1) + V(L2) + V(L3) + V(L4) + V(L5). The program runs down the stored list and accepts the first word for which VAL(N) = HSUM (N-1).

There are two basic ways to use this algorithm. In the first, the program makes a guess more or less at random and then proceeds down the VAL = HSUM path. In the second, the program uses four or five preselected guess words, records the hits, but makes no use of this information until it goes to make the fifth or sixth guess. In either procedure, a few words are placed at the head of the main list for various reasons. The rest of the list is scrambled.

A major weakness of Jotto is that anagrams are indistinguishable from one another. Once an anagram of the target is found, there is no alternative to blindly guessing other anagrams until the opponent says "that's it". I mitigate the problem somewhat by placing one member of each large set of anagrams at the head of the list. The largest OSPD set is AEPRS with eleven members. If one of its members (an anagram of the target) is hit with the fifth guess, we can expect to hit the target by the tenth. The worst case will be 15 guesses. The locations of the anagrams in the list are scrambled after each game.

Words containing AERS and one other letter are the worst targets in Jotto. Table 1 lists 93 of them; the fifth letter can be every one except Q. Should the program start guessing with BREAD 3, CARES 4, it can get trapped in a long path something like FEARS 4, RAGES 4, SHEAR 4, ARISE 4, RAJES 4, ... RAZES 5. This is where starting with four or five non-overlapping words comes into its own. Guessing with a bunch of consonants early on helps avoid the 4-4-4 trap. Both of my Jotto programs start with a word list that was scrambled when first made. In the pure VAL = HSUM method, this list is further scrambled each time it is read in. This helps prevent an opponent from taking advantage of the AERS? and similar weaknesses. (The programs go through a few other shenanigans not worth discussing.)

The article "A New Look at the Jotto Problem" in the November issue abstracts a very extensive list of five-word starter combos for Jotto. These combos were found by an iterative process. They can not be developed in the course of a Jotto game,as some may imply from the article. None use all OSPD words, but four non-overlapping words (all OSPD) serve well enough for our game. Table 2 gives a set of five-word combos with OSPD words, but only the first four words are non-overlapping (use 20 different letters) and have one vowel per word; the fifth word contains the remaining two vowels and has one consonant in common with the others. These words are used as follows: Choose a combo at random from several included in the program. Take the first three members as guesses. If they make five hits on the target, skip the fourth member and continue the search with the VAL = HSUM method. Otherwise, take the fourth word as a guess and then continue with VAL = HSUM. But, if there are less than four hits with the first four words, take the reserve word as the fifth guess. The seven-guess finding for KALIF (below) is a perfect illustration of its use.

For target words chosen at random, the pure VAL = HSUM and the present combo methods do about equally well, with an average of 8.2 guesses per game. I prefer the second method for the reason given above. Here are sample games using the second method:

target	guesses needed	average
QUITE	8, 5, 8, 8, 7, 8, 6, 9, 8, 6	7.3
EMPTY	11, 9, 8, 8, 7, 6, 8, 7, 7, 8	7.9
RAZES	9, 6, 7, 7, 9, 8, 8, 7, 9, 9	7.9
SCHAV	8, 8, 8, 8, 8, 9, 7, 8, 8, 9	8.1
KALIF	9, 8, 8, 7, 9,10, 7,10, 9, 8	8.5
ENVOI	7, 8, 8, 9, 9, 8, 7, 8, 8, 9	9.1
ESTOP	6, 9, 7,11, 9, 6, 8, 8, 9,10	9.1
BREAD	9,10,12, 8, 8, 6, 8,10,12,10	9.3

gawks 1, vetch 1, fjord 2, plumb 1, torus 1, mowed 2, acold 2, pared 4, debar 5, ardeb 5, beard 5, bread 5
gambs 1, wench 1, fjord 0, klutz 1, privy 2, spiel 2, inapt 2, pubic 1, lying 1, peaky 3, empty 5
dwarf 2, blitz 1, jocks 1, squeg 2, tears 4, arise 4, laser 4, sabre 4, razes 5
bortz 0, vangs 1, fjeld 2, whump 0, kylix 3, yelks 2, kalif 5
rewax 1, blitz 0, fjord 0, gucks 2, nymph 1, quays 2, vangs 3, nevus 2, schav 5
waltz 1, gecks 1, fjord 0, squib 3, quite 5

THE CRASH ALGORITHM

Books by Tony Augarde, Gyles Brandreth, and David Parlett all mention (usually only briefly) various versions of Giotto. The game that I like best is often called Crash; that title was used by Dave Silverman in an August 1969 Word Ways Kickshaws. The best description of the game is in Peter Newby's *Pears Word Games* which calls it Words. There is also a boxed version called Words Worth.

After describing Words, Peter Newby says "Jotto is logical deduction for the hard of thinking". Words is the better game, but it is not that much different from Jotto. In Words (= Crash), a player reports a hit if the guess word has a letter corresponding to one in the same position in the target; for example, BearD scores two hits on BreaD. Newby's procedure for guessing the target can be improved upon. The algorithm is identical to that described above for Jotto except we keep a separate score for each letter position.

Newby refers to E as the most commonly used letter in the alphabet. For our purposes this is not correct. Let's look at the data. Table 3 lists occurrences of letters in various positions for five-letter OSPD words. For isograms, E, S, and A are about equally frequent. I didn't use these frequencies for choosing starter words for Jotto, but I did

use them for playing Crash with isograms (discussed in the next section). For playing Crash with the full OSPD list, we must consider frequencies in each of the five positions separately. E is hardly used in first position and S is hardly used in second. Let's use those facts to establish starting words for Crash.

In each of the following combos, there are five different letters in each position (column). The words are chosen to use the most frequent letters possible. The left combo is almost perfect in that respect. I am not sure that a statistical approach is best, but I use these in Crash for lack of any better idea. Of the 8184 OSPD words, only 135 do not get hit by the first combo. The other combos are not quite as good but serve to add variety and concealment to the guessing. If we add ARENA, INDOL and FLUSH to the left combo, no OSPD word gets by, but using more than five words in a starter set leads to diminishing returns.

peris	tunas	brats	delis
cooed	cried	toned	boned
talar	salty	saint	carat
suite	aerie	curie	slaty
bialy	boart	delay	thine

Like Jotto, Crash has a weakness. There are 134 OSPD words of the form ?A?ES, given in Table 4. The fix for Crash is different than for Jotto. Here we have to rely on catching the problem early and jumping out of the starter set. Whenever there are three or more hits in the starter words, the program switches to the VAL = HSUM sequence. Note that ?A?ES will always be detected in three guesses with the above starts. Here are some test reports. The average number of guesses seems to be about 10.

target	guesses	average
WALTZ	5, 6, 8, 7,10, 8, 7, 9, 6, 6	7.2
QUITE	8, 6, 5, 7,11, 6, 7, 7, 8, 8	7.3
BREAD	8, 9, 7, 9, 7, 5, 7, 9, 6, 6	7.3
KALIF	8, 7, 8, 8, 7,10, 6, 8, 8, 9	7.9
EMPTY	8,10,10, 7, 9,10,10, 9,10, 9	9.2
RAZES	12, 8,11, 8,12,13, 7, 6, 9,13	9.9
DESEX	9,11,11, 9,10,12,10,10, 8,11	10.1
SCHAV	11,10, 8, 9, 9,12,11,11,10,13	10.4
HEIGH	11,12,10, 9,11,11, 8,11,12,10	10.5
VROOM	13, 7, 9,11,10,11,10,12,10,13	10.6
QUAFF	11,12,11, 9,10,11,11,12, 9,12	10.8
ESTOP	10,12,10,10,13,12,10,13,10, 9	10.9
ENVOI	10,11,11,12,12,11,10,10,11,12	11.0
JUJUS	13,13,13,13,10,13,12,10,11,10	11.8

peris 0, cooed 1, talar 1, suite 0, bialy 1, cabby 0, liner 0, bread 5
delis 0, boned 0, carat 0, slaty 1, thine 2, stone 1, shill 1, alike 2,

plink 1, unite 3, quite 5
brats 1, toned 1, saint 1, fates 3, cafes 3, eaves 3, wames 3, mares 3,
rapes 4, raxes 4, rages 4, races 4, razes 5
delis 0, boned 0, carat 0, slaty 0, thine 0, uteri 0, achoo 1, amuck 0,
oxbow 1, mucor 1, gipon 1, proof 3, vroom 5
peris 0, cooed 0, talar 0, suite 1, bialy 1, skull 0, disme 0, wrath 1,
quasi 3, quart 3, quack 3, quaff 5

A fascinating aspect of playing with either the Jotto or the Crash programs is to watch how they sometimes find a target without having made many early hits. ENVOI and ESTOP do not crash with any of the above starter combos. Here are some guessing sequences that find them:

peris 0, cooed 0, talar 0, suite 0, bialy 0, glyph 0, intro 1, uteri 1,
ennui 3, envoi 5
tunas 0, cried 0, salty 0, aerie 0, boart 0, rheum 0, lymph 0, oxbow 1,
indol 2, nicol 1, envoi 5
delis 0, boned 0, carat 0, slaty 0, thine 0, octyl 0, edema 1, wreck 0,
hydra 0, eikon 2, ephor 2, envoi 5
delis 0, boned 0, carat 0, slaty 0, thine 0, fresh 0, oxbow 1, extra 0,
estop 5
delis 0, boned 0, carat 0, slaty 0, thine 0, hyrda 0, fresh 0, achoo 1,
awful 0, oxbow 1, mucor 1, piton 2, estop 5

THE ISOGRAM CRASH ALGORITHM

Here are starter combos for Isogram Crash (I-Crash). 88 words are not hit by the first combo. (The following eight words hit all of the isograms: CRATE, DELAY, FIORD, PURIN, SAINT, TONER, ALTOS, THECA.)

bolas	cions	dairy	pries
tined	saned	tines	carte
saint	boite	crate	solar
curie	train	serin	aloin
peaty	deray	bolar	diary

Here are results of a few tests:

target	guesses	average
BREAD	5, 6, 8, 5, 8, 8, 8, 6, 8, 7	6.9
QUITE	10, 8, 6, 5,10, 7, 7, 8, 8, 8	7.7
EMPTY	8,10,10,10,10, 9,10, 9, 7,11	9.4
KALIF	9,11, 9,10, 8, 8,12,10,10, 8	9.5
RAZES	11,10,10, 8, 8, 7,10,11,10,12	9.7
SCHAV	9, 8, 9,11,10, 9,11,10,11,10	9.8

COMPARING THE THREE GAMES

Here is a very brief comparison of the three guessing games. About all we can say here is that Jotto is a bit faster than Crash.They are two different games so this is not important. BREAD is an easy target for Crash but gets fouled by the anagram problem in Jotto. ESTOP also contains frequently used letters but gets fouled in the anagram trap. It is a difficult target in both Jotto and Crash, but for different reasons.

	JOTTO		CRASH		I-CRASH	
	avg	range	avg	range	avg	range
BREAD	9.3	6 to 12	7.3	5 to 9	6.9	5 to 8
EMPTY	7.9	6 to 11	9.2	7 to 10	9.4	7 to 11
KALIF	8.5	7 to 10	7.9	6 to 10	9.5	8 to 12
QUITE	7.3	6 to 9	7.3	5 to 11	7.7	5 to 10
RAZES	7.9	6 to 9	9.9	6 to 13	9.7	7 to 12
SCHAV	8.1	7 to 9	10.4	8 to 13	9.8	8 to 11
ESTOP	9.1	6 to 11	10.9	9 to 13		
ENVOI	9.1	7 to 9	11.0	10 to 12		

GENERAL OBSERVATIONS

A Giotto-style commercial game called Mastermind was popular in 1975. It used six different colored pegs to define four-peg "words". Repeated colors in the target was allowed. Later, a five-peg version using eight colors was produced. Guessing was a combination of Jotto and Crash rules. Mastermind can be played with words instead of colored pegs. For example, if the target is STAIR and the guess is SOUTH, the report would be "one hit (S) and one letter out of place (T)". If the guess were SOUSE, the report would be "one hit" because the S in STAIR is only counted once. If the guess were TARRY, the response would be "three letters out of place". Only one R is counted. This game would probably be best with isograms to avoid the complicated rules. I may program this game in the future. One of the above mentioned books describes Jotto as "Word-Mastermind".

Two players can use my one-sided computer programs in a competitive game as follows: Player A picks a target and asks Player B to guess it. At the same time, Player A has the computer guess it. Player A scores the number of guesses the computer needed plus or minus the difference between the number of guesses Player B needed compared to the computer. This introduces a curious aspect to Jotto or Crash. Most players would have a tendency to pick obscure words for their opponent. The computer does well with these words. On the other hand, if a player picks one of the words (discussed above) which are tough on the computer, his human opponent might guess it easily.

I used the OSPD for this study mainly because it was available on disk. I would have preferred to use Merriam-Webster's 10th Collegiate, because the OSPD is too inclusive for games (including Scrabble). For

logology (or Scrabble tricks), we must use the largest database which is both available and manageable. Game playing is different. Giotto games with six-letter words should be interesting, but here too I would prefer to limit the vocabulary to 10th Collegiate.

TABLE 1

BARES baser bears braes saber sabre ACRES escar cares carse races scare serac DARES dears rased reads FARES fears safer AGERS gears rages sager sarge HARES hears rheas share shear ARISE raise serai RAJES ASKER eskar rakes saker ARLES earls lares laser lears rales reals seral MARES marse maser reams smear EARNS nares nears saner snare AROSE APERS asper pares parse pears prase presa rapes reaps spare spear AURES urase ureas ursae AVERS raves saver WARES wears resaw sawer sewar sware swear RAXES EYRAS resay sayer years RAZES

TABLE 3

	ISOGRAMS						ALL 5-LETTER WORDS					
Q	42	9	6	0	0	57	Q 51	11	8	0	0	70
J	85	3	11	7	0	106	J 126	6	17	9	0	158
Z	30	12	45	28	10	125	X 12	40	85	7	41	185
X	7	28	65	5	31	136	Z 41	14	75	59	14	203
V	113	20	120	57	2	312	V 157	38	160	91	2	448
F	314	8	49	59	20	450	W 285	96	149	87	39	656
W	218	58	119	70	28	493	F 414	14	120	145	49	742
K	102	24	106	223	131	586	K 180	52	129	295	174	830
B	389	25	123	65	20	622	B 566	50	200	130	31	977
G	267	21	139	170	74	671	G 389	46	225	253	83	996
Y	66	111	73	45	489	784	H 319	360	64	126	243	1112
H	222	269	47	96	182	816	M 419	114	319	255	119	1226
M	281	74	208	171	88	822	Y 81	153	114	59	831	1238
P	374	99	140	157	85	855	P 531	146	229	259	113	1278
C	456	80	183	194	72	985	C 639	118	249	276	93	1375
D	289	31	145	195	431	1091	U 108	748	426	240	25	1547
U	83	501	330	179	20	1113	D 424	56	259	291	565	1595
N	107	149	423	377	264	1320	N 176	232	591	506	347	1852
T	344	99	253	369	351	1416	T 527	165	404	592	459	2147
L	266	347	341	292	195	1441	L 380	488	555	516	336	2275
O	122	822	365	234	113	1656	I 112	908	716	544	121	2401
I	68	611	541	408	62	1690	O 176	1255	645	417	194	2687
R	242	444	565	306	328	1885	R 388	618	751	447	505	2709
A	269	968	617	391	156	2401	A 474	1396	823	611	334	3638
S	562	36	127	199	1647	2571	E 188	999	548	1635	897	4267
E	84	553	261	1105	603	2606	S 1021	61	323	334	2569	4308

TABLE 2

black vents fjord whump zingy qx	blawn kempt fjord vughs cylix qz
blawn kempt fjord vughs zincy qx	brawn fjeld compt vughs kylix qz
brawn fjeld compt vughs zinky qx	champ fjeld bortz swung kylix qv
gambs wench fjord klutz privy qx	gambs wench fjord klutz pyxis qv
gawks fjeld bortz chump vinyl qx	gawks fjeld bortz chump vying qx
gawks fjeld bortz munch privy qx	gawks fjeld bortz munch pyxis qv
gawks fjeld bortz nymph quick vx	gawks vetch fjord plumb zincy qx
gawks vetch fjord plumb zingy qx	gawks vetch fjord plumb zinky qx
gnaws fjeld bortz chump kylix qv	gnaws vetch fjord plumb kylix qz
gnaws vetch fjord plumb zinky qx	jacks veldt bring whomp furzy qx
snack fjeld bortz whump vying qx	swamp bench fjord klutz vying qx
swamp fjeld bortz chunk vying qx	swang fjeld bortz chump kylix qv
swang vetch fjord plumb kylix qz	swang vetch fjord plumb zinky qx
swank fjeld bortz chump vying qx	swank vetch fjord plumb zingy qx
twang fjeld crumb qophs kylix vz	twang fjeld crumb qophs zinky vx
vamps bench fjord klutz wingy qx	vamps fjeld bortz chunk wingy qx
vamps fjeld botch wrung kylix qz	vamps fjeld botch wrung zinky qx
vamps fjeld butch grown kylix qz	vamps fjeld butch grown zinky qx
vamps fjeld butch wrong kylix qz	vamps fjeld butch wrong zinky qx
vamps wench fjord klutz bigly qx	vamps wench fjord klutz bilgy qx
vangs fjeld bortz chump kylix qw	vangs fjeld bortz whump cylix kq
vangs fjeld bortz whump dicky qx	vangs fjeld bortz whump kylix cq
vangs fjeld bortz whump picky qx	vangs fjeld bortz whump quick xy
vangs fjeld brick whump zloty qx	vangs fjeld brock whump ritzy qx
vangs fjeld brock whump sixty qz	vangs fjeld brock whump xysti qz
vangs wecht fjord plumb kylix qz	vangs wecht fjord plumb zinky qx
waltz gecks fjord nymph squib vx	waltz kemps bunch fjord vying qx
waltz kench bumps fjord vying qx	whack vents fjord plumb zingy qx
whang fjeld bortz mucks privy qx	whang fjeld bortz mucks pyxis qv
whang fjeld bortz skimp curvy qx	whang kemps blitz fjord curvy qx
wrack fjeld thumb zings proxy qv	dwarf blitz jocks nymph squeg vx
jacks bortz fling whump gyved qx	vangs blitz fjord whump pecky qx
wacks blitz fjord nymph squeg vx	whack bortz fling jumps gyved qx
whang blitz frock jumps gyved qx	whang blitz frump jocks gyved qx
whang bortz flick jumps gyved qx	gecks blitz fjord nymph squaw vx
vetch fjord plumb zings gawky qx	vetch fjord plumb zings pawky qx
blitz fjord gucks nymph dewax qv	blitz fjord gucks nymph rewax qv

TABLE 4

BaBes BaKes BL BN BR BS BT CD CF CG CK CM CN CP CR CS CT CV
DaCes DaLes DM DR DT DZ ES EV FC FD FK FM FN FR FT FX FZ
GaGes GaLes GM GP GS GT GZ HD HJ HK HL HM HR HT HV HZ JD JK JN JP
KaLes KaMes KN LC LD LK LM LN LR LS LV LZ MC MG MK ML MN MR MT
MaZes NaMes NP NR NT NV OS OV PC PG PL PN PR PS PT PV PX RC RG
RaJes RaKes RL RP RS RT RV RX RZ SB SD SF SG SK SL SN ST SV SX
TaBes TaCes TJ TK TL TM TP TR TT TX VL VN VS WD WG WK WL WM WN
WaNes WaRes WV WX ZX