ABBREVIATIONS WITHOUT AMBIGUITIES

A. ROSS ECKLER
Morristown, New Jersey

Suppose that one wishes to replace the words in a commonly-used list by abbreviations in order to save space in texts. How should the abbreviations be chosen in order to avoid confusion? At first blush, it would appear sufficient to ensure that each abbreviation is distinct from the others. But this is not enough; one should also require that (1) the abbreviation be formed according to specified rules, and (2) under those rules, the abbreviation could have been formed from only one member of the list. Thus, a human (or a computer) looking at an abbreviation, knowing the list of words and the rules, can unambiguously determine the word in the list that the abbreviation stands for; it is no longer necessary to memorize the abbreviation corresponding to each word, or have a table of these at one's elbow.

To fix ideas, consider a well-known word list with abbreviations: the US Post Office abbreviations for the 50 states. Although each abbreviation is different from its mates, one cannot unambiguously reconstruct state names from abbreviations. The two rules that the Post Office appears to have followed are (1) first letter of abbreviation = first letter of state name, (2) second letter of abbreviation = first letter of second word in two-word state name. Beyond this, it is impossible to discern consistent rules. Setting aside the complicated cases of states beginning with A or M, one has several possibilities for choosing the second letter of the abbreviation of a single-word state: (3) take final letter of state name if all final letters are different, (4) take second letter of state name if all second letters are different, (5) if the second letters are the same, take the third letters instead. The Post Office seems to have adopted rule (3) for states beginning with C,D,G, H,K,L,P and V, rule (4) for states beginning with D,F,I,O and W, and rule (5) for states beginning with N (NV for Nevada, NB the original choice for Nebraska) and T (TN for Tennessee, TX for Texas). Note that Delaware qualifies for either rule (3) or rule (4), as does California, Mississippi, Missouri and Tennessee.

Using rules (1) and (2) alone to determine state name abbreviations, one finds that decoding abbreviations is riddled with ambiguities. For example, AL = Alabama could be also decoded as Alaska, and AK = Alaska could equally well be the abbreviation for Arkansas. The table on the next page summarizes the 16 states having a total of 31 ambiguities.

Could the state name abbreviations have been designed to yield fewer ambiguities? The answer to this question is yes; three such

Abbr	Could also stand for
AL	Alaska
AK	Arkansas
AR	Arizona
CO	California
ID	Indiana
ME	Massachusetts, Minnesota
MA	Maine, Maryland, Montana, Michigan, Minnesota
MI	Maine, Minnesota, Mississippi, Missouri
MN	Maine, Michigan, Montana, Maryland
MS	Massachusetts, Missouri
MO	Minnesota, Montana
MT	Massachusetts, Minnesota
NE	Nevada
ND	Nevada
OH	Oklahoma
WI	Washington, Wyoming

rules are defined and evaluated below. The three rules are:

Rule (a) code the 10 two-word states with the first letters of their two words, and the remaing 40 with their first letter followed by any other letter in the name

Rule (b) code the 10 two-word states with the first letters of their two words, and the remaining 40 with any bigram taken from the name

Rule (c) code the 10 two-word states with the first letters of their two words, and the remaining 40 with any two letters in their name, in the proper order

When decoding an abbreviation, first check it for matches against the two-word states; if no match is found, then use the single-word state encoding rule to decide which state it could be.

The table below indicates possible choices for unambiguous abbreviations for the 40 one-word state names, or (in brackets) choices that minimize the ambiguity among all abbreviations available for that state name, along with the states it is confused with. For example, the notation [AL=AL; AS,AK=AR] means that the abbreviation AL can be decoded as Alabama as well as Alaska using Rule (a), and that the abbreviations AS and AK can be decoded as Arkansas as well as Alaska using Rule (a). Furthermore, no other possible choices for an abbreviation using Rule (a) do this well; for example, AA can be mis-decoded as Alabama, Arizona or Arkansas.

	Rule (a)	Rule (b)	Rule (c)
AL	AB,AM	AB,BA,AM	LB,AB,BM
AK	[AL=AL;AS,AK=AR]	[SK=NE]	LK
AZ	AZ,AI,AO	IZ,ZO	AZ,RZ,IZ,ZO,ZN,ZA
AR	[AK,AS=AK;AR=AZ]	RK	[RK,RS=NE[
CA	CA,CF	IF,FO,RN	AF,CF,LF,FN,IF
CO	CD	DO,OL	CD
CT	XT,CC,CE	EC,CT,TI,CU	CC,OC

DE	DE,DL,DA,DW,DR	DE,EL	LE,DE,DW,LW,DL,WR,DR
FL	FL,FO,FR,FD,FA	FL	FL,FD
GA	GE,GO,GR,GI,GA	GE,EO	GE,GG,GR
HI	HA,HW,HI	HA,II	HW
ID	IH	[HO=OK;ID=FL]	DH
IL	IL,IS	IL,LL,NO,OI	IL
IN	[IN,II=IL;ID=ID]	DI	DN
IO	IW	OW	IW,OW
KS	KA,KS	[SA=AR,MA;NS=AR,PN]	[KS=AR]
KY	KE,KT,KU,KC,KY	KE,EN,TU,UC,CK,KY	KC,KK,KY,UK,TK,CK,KU
LA	LO,LU,LI,LS,LN,LA	UI	LU,UN
ME	[ME=MA,MN]	[AI=HI]	[IE=MN;ME=MA or MN]
MD	MD,ML,MY	RY,ND	MD,MY
MA	[MH,MC=MI;MU=MO;MT=MN]	AC,HU,US,ET,TT,TS	UE,HE,HS,AU,HU,AC
MI	MG	IG,GA	CG
MN	[MT=MT,MA;MO=MT,MO]	OT	[IE=ME;ME=MA or ME]
MS	MP	IP,PP,PI	SP,IP,PP,MP
MO	[MU=MA;MR=MD]	UR	UR
MT	[MT=MN,MA;MO=MO,MN]	[MO=VT;NT=KY]	[TN=TN,WA]
NE	NB,NS,NK	EB,BR	NB,EB,BR,BS,BK
NV	NV	EV	VD
OH	OI	OH	[OH=OK]
OK	OK,OM,OL,OA	OK,KL	KL,KM,KH,LH,HM
OR	OR,OE,OG,ON	EG,GO	OE
PA	PE,PN,PS,PY,PL,PV,PI,PA	PE,SY,LV	PA,PE,PN,PY,PL,PV,PA,YV,LV
TN	TN	EE	EE SY,SV
TX	TX,TA	EX,XA	TX,EX,XA,XS
UT	UT,UA,UH	[UT=CT;AH=OK]	UH,TH
VA	VI,VG,VA	VI,IR	VG
VT	VE,VM,VO,VT	VE,ER,RM	VE,VM,VO,VT
WA	WA,WH,WT	TO,GT	AG,WT,SG,HG,GT
WI	WC	WI,SC	WC
WY	WY,WM	WY,YO	WY,WM,YM,YG

Rule (a) generates 8 ambiguous states with a total of 11 different alternatives; rule (b) generates 6 ambiguous states with a total of 7 different alternatives; rule (c) generates 6 ambiguous states with a total of 8 different alternatives. Maine and Montana are always ambiguous. Note that Minnesota and Maine must share the abbreviations IE and ME, with one and two ambiguities respectively, under rule (c).

A still-unanswered question: does there exist a set of rules leading to fewer ambiguities than the ones above have? The ideal, of course, would be a set of rules for which no ambiguities exist, but this seems unlikely. One must probably go to three-letter abbreviations to achieve this desideratum.