

## Inter-Office Memorandum

To	Whom it May Concern	Date	July 19, 1980
From	Lyle Ramshaw	Location	Palo Alto
Subject	Special characters over the years	Organization	CSL

# XEROX

Filed on: [Maxc1]<Fonts>SpecialCharacters.bravo  
[Maxc1]<Fonts>SpecialCharacters.press  
[Ivy]<Fonts>Memos>SpecialCharacters.bravo  
[Ivy]<Fonts>Memos>SpecialCharacters.press

This memo discusses the allocation of character codes over the years in the fonts TimesRoman, Helvetica, and Math, and also deals with some related issues pursuant to the July 1980 font cataclysm. In this memo, I will merely list the assignments of the special character codes. Since I want the reader to be able to print this memo before the cataclysm, I was not able to include any demonstration of what the new special characters of the third generation actually look like.

Since the cataclysm has already occurred, there is a new document on [Ivy]<Fonts> called CloverCharacters.Press. This is a sort of companion to the documentation file CloverFonts.Press on the same directory. Where CloverFonts simply lists the names, sizes, and faces of the fonts on Clover, CloverCharacters has a one page table for each family that shows what printing character is associated with each character code. CloverCharacters only discusses the post-cataclysm truth, and hence doesn't give the sense of change and history provided by this memo. On the other hand, CloverCharacters has pictures instead of words, and CloverCharacters includes all of the families on Clover, not just TimesRoman, Helvetica, and Math.

### **TimesRoman and Helvetica:**

Let's ignore Math for the present, and concentrate on TimesRoman and Helvetica. Almost every Roman font agrees on the meanings of almost all of the character codes from #40 through #176. These character codes are defined in PARC Roman fonts by the following rough equivalent of ASCII:

- #40=space <see discussion below>
- #41=exclamation point
- #42=symmetric double quote
- #43=number sign, sharp sign, pound sign
- #44=dollar sign
- #45=per cent sign
- #46=ampersand
- #47=single quote <see discussion below>
- #50=left parenthesis
- #51=right parenthesis
- #52=asterisk
- #53=plus sign
- #54=comma
- #55=minus sign or hyphen <see discussion below>
- #56=period
- #57=slash

#60 through #71=digits 0 through 9  
 #72=colon  
 #73=semicolon  
 #74=left angle bracket (*not* less than sign)  
 #75=equal sign  
 #76=right angle bracket (*not* greater than sign)  
 #77=question mark  
 #100=at sign  
 #101 through #132=upper case A through Z  
 #133=left square bracket  
 #134=backslash  
 #135=right square bracket  
 #136=up arrow  
 #137=left arrow  
 #140=<see discussion below>  
 #141 through #172=lower case a through z  
 #173=left curly bracket  
 #174=vertical bar  
 #175=right curly bracket  
 #176=tilde symbol (*not* accent).

The history of character sets at PARC has mostly involved the fate of the other ASCII character codes.

### First Generation:

In the first generation fonts:

#40 was the only space there was  
 #47 was a symmetric single quote  
 #55 was a minus sign  
 #140 was not defined.

In addition, there was one special character: #30=<sup>^</sup>X was defined to be an underline accent character. Gacha is an example of a first generation font that has survived to the current day.

### Second Generation Changes:

The second generation added nineteen special characters, and affected a few of the basic codes as well. The changes to the basic characters were:

#40 became a space designed to go between words  
 #47 became a closing single quote  
 #55 became a hyphen  
 #140 was still undefined.

The special characters introduced were:

#2=<sup>^</sup>B=upsidedown question mark  
 #3=<sup>^</sup>C=a cedilla accent on a lower case c  
 #4=<sup>^</sup>D=umlaut accent  
 #5=<sup>^</sup>E=grave accent  
 #6=<sup>^</sup>F=ff ligature  
 #7=<sup>^</sup>G=opening single quote  
 #10=<sup>^</sup>H=upsidedown exclamation point  
 #13=<sup>^</sup>K=acute accent  
 #16=<sup>^</sup>N=minus sign  
 #17=<sup>^</sup>O=em quad  
 #20=<sup>^</sup>P=tilde accent  
 #21=<sup>^</sup>Q=ffi ligature

#22=^R=ffl ligature  
 #23=^S=em dash  
 #24=^T=fi ligature  
 #25=^U=fl ligature  
 #26=^V=en dash  
 #30=^X=minus sign, duplicating #16  
 #31=^Y=figure space, a space just as wide as a digit  
 #34=^\  
=en quad.

### Third generation:

In the third generation release, the region of character code space from #1 through #37 was pretty much filled up, and extra characters began to find their way into the eight bit character codes, those between #200 and #377. The changes to the basic codes were:

#40 remained a word space  
 #47 remained a closing single quote  
 #55 remained a hyphen  
 #140 now duplicates #7, an opening single quote.

This last change was an excellent choice from a public relations point of view, since much of the Arpa community was already committed to this convention.

In third generation fonts, the first 40 character codes are allocated as follows:

#0=ASCII null  
 #1=^A=hacek accent, upsidedown circumflex  
 #2 through #10 are as in second generation  
 #11=^I=ASCII tab, used by Bravo  
 #12=^J=ASCII line feed, treated specially by Bravo  
 #13=^K=acute accent, as in second generation  
 #14=^L=ASCII form feed, page break in Bravo  
 #15=^M=ASCII carriage return, used by Bravo  
 #16=^N=macron accent (!Unlike second generation!)  
 #17 through #26 are as in second generation  
 #27=^W=breve accent  
 #30=^X=minus sign (no longer duplicated at #16)  
 #31=^Y=figure space, as in second generation  
 #32=^Z=ASCII end-of-file and Bravo paragraph break  
 #33=^[\_=ASCII Escape, used by Bravo  
 #34=^\  
=en quad, as in second generation  
 #35=^]=low dot accent  
 #36=^^=duplicate of tilde accent at #20, since #20  
       has been stolen by some Bravo's for something  
 #37=^\_ =circle accent.

The change of #16=^N from one of the minus signs to a macron accent is the only incompatible feature of the change from the second to the third generation.

As the previous listing indicates, the seven bit codes are now pretty well filled up; from #40 through #176 are the basic codes discussed at the beginning, and #177 is ASCII delete. Large-scale expansion of character sets must therefore either involve introducing new fonts or giving printing meanings to eight bit character codes. Here is the list of the eight bit codes that Pellar choose to define in the third generation fonts:

#200=umlaut on an upper case A  
 #201=umlaut on an upper case O  
 #202=circle accent on an upper case A  
 #220=ff ligature

#221=ffi ligature  
 #222=ffl ligature  
 #223=fi ligature  
 #224=fl ligature  
 #230=en quad  
 #231=em quad  
 #232=figure space  
 #233=en dash  
 #234=em dash  
 #235=minus sign  
 #236=en leader  
 #237=5-em space  
 #240=umlaut on a lower case a  
 #241=umlaut on a lower case o  
 #242=circle accent on a lower case a  
 #243=umlaut on a lower case u  
 #244=cedilla accent on a lower case c  
 #250=umlaut accent  
 #251=circle accent  
 #252=grave accent  
 #253=acute accent  
 #254=circumflex accent  
 #255=tilde accent  
 #256=breve accent  
 #257=macron accent  
 #260=cent sign  
 #261=sterling sign  
 #265=star  
 #266=section sign  
 #267=bullet  
 #270=dagger  
 #271=double dagger  
 #272=paragraph sign  
 #274=plus-minus sign  
 #275=upside-down question mark  
 #276=upside-down exclamation point  
 #277=underscore  
 #337=non-required hyphen  
 #350=hacek accent  
 #351=low dot accent

By the way, the accents associate with the #250 through #257 and #350 and #351 codes are not zero width; they should only be used by programs that are smart enough to center them over the accented character.

#### **Some thoughts about the eight bit characters:**

From the CSL point of view, it is rather unfortunate that interesting characters are now beginning to be associated with eight bit character codes. Much of CSL's software world was written with only seven bit character codes in mind, and probably won't be changed in the near future. Consider Tex for example: this document compiler is used by a community of people at PARC. But Tex was written on a PDP-10, and has the assumption that character codes are seven bits long burned into its guts. This means that whatever pretty new characters Pellar sticks up in the eight bit region will probably never be accessible from Tex.

Above and beyond this long term worry, these eight bit characters have affected the plan for the font cataclysm. As described in the AltoFontGuide memo, I chose to strip the eight bit characters out of the standard Alto fonts. In addition, there are two different Fonts.Widths files for use after the cataclysm: the standard one ([Ivy]<Fonts>Fonts.Widths) has the eight bit characters stripped out, while the other ([Ivy]<Fonts>EightBit>Fonts.Widths) has them left in. Producing this stripped-down width dictionary demanded a little hacking on my part, but consider the alternative. The only other choice would have been to go over every program in the CSL software world that reads Fonts.Widths, and check that they can correctly discard width information about characters with eight bit character codes. And just consider the list of affected programs: Bravo 7.5, SIL, PressEdit, Laurel even Pub! The complexity of having two different width dictionaries around pales into insignificance before the unpleasantness of hacking on Pub.

On the other hand, I plan to leave all of the eight bit characters in the dictionary of printing fonts. The printing servers and such font systems as Prepress have already been checked out on eight bit characters, and handle them correctly..

### Math:

We bring this memo to a close by recounting the sad history of the Math font. The first generation of the Math font had the character map:

- #40=printing symbol for a space
- #41=dagger
- #42=degree symbol
- #43=infinity symbol, lazy eight
- #44=cent sign
- #45=division sign, elementary school version
- #46=logical and, an A without the bar, lattice meet
- #47=equal sign with dot above
- #50=<unassigned>
- #51=<unassigned>
- #52=high dot
- #53=plus-or-minus sign
- #54=contains as an element, such that
- #55=minus-or-plus sign
- #56=three dots meaning therefore
- #57=slash in a circle
- #60=circle
- #61=box
- #62=triangle with flat side on the bottom
- #63=diamond
- #64=plus in a circle
- #65=minus in a circle
- #66=multiplication x in a circle
- #67=angle sign
- #70=star
- #71=high dot, darker than #52
- #72=section sign
- #73=black slug
- #74=less than or equal sign
- #75=not equal sign
- #76=greater than or equal sign
- #77=upsidedown question mark
- #100=in care of
- #101=for all, upsidedown A

#102=is an element of  
#103=blackboard boldface C, the complex numbers  
#104=nabla, upsidewind upper case Greek delta  
#105=there exists, backwards E  
#106=double dagger  
#107=is a proper subset of  
#110=is not a proper subset of  
#111=is a subset of  
#112=contains as a proper subset  
#113=does not contain as a proper subset  
#114=contains as a subset  
#115=printing symbol for carriage return  
#116=is not an element of  
#117=the null set  
#120=is proportional to  
#121=right and left arrows with right on top, reversible reaction  
#122=blackboard boldface R, the real numbers  
#123=wavy equals  
#124=is perpendicular to  
#125=union  
#126=logical or, a V shape, lattice join  
#127=three bar equal  
#130=a multiplication x  
#131=arrow starting out rightwards, then pointing down  
#132=solid triangle pointing rightward  
#133=curly left angle bracket  
#134=thick slash  
#135=curly right angle bracket  
#136=down arrow  
#137=right arrow  
#140=<unused>  
#141=aleph  
#142=wavy right arrow  
#143=copyright symbol, small C in a circle  
#144=partial sign  
#145=equal sign with top bar wavy  
#146=double-headed arrow  
#147=double shafted right arrow  
#150=slashed h, for Planck's constant  
#151=left ceiling bracket  
#152=right ceiling bracket  
#153=left floor bracket  
#154=right floor bracket  
#155=double vertical bar  
#156=logical negation sign  
#157=small circle operator, composition  
#160=T on its side, vertical bar on the left  
#161=double shafted T on its side, vertical bar on the left  
#162=registered trademark symbol, small R in a circle  
#163=northeast arrow  
#164=northwest arrow  
#165=southwest arrow  
#166=southeast arrow  
#167=much less than sign

#170=much greater than sign  
 #171=intersection  
 #172=T on its side, vertical bar on the right  
 #173=the fraction 1/4  
 #174=the fraction 1/2  
 #175=the fraction 3/4  
 #176=right and left arrows with one-sided heads, left on top.

When Ron Pellar prepared the second generation of Math, he added a few new characters:

#2=^B=less than sign  
 #3=^C=greater than sign  
 #4=^D=a one point space, for use in equations  
 #50=a prime, thin enough to be doubled up.  
 #51=upper case Greek pi, to reduce the need for Hippo.

In addition to making these additions, Ron decided to try replacing thirteen of the original characters with more useful printing roles. This incompatibility generated such storms of protest that Ron decided to rescind these replacements in the third generation of Math. If you really want to know what these characters were, you should look at the November 1978 or September 1979 edition of the Alto User's Handbook, the Bravo font page. Unfortunately, that page was prepared using second generation Math. I am not going to mention those replacement characters here, because it is my fond hope that all memory of them should vanish from off the face of the earth.

On to third generation Math! The following list describes the character code assignments of third generation Math where they differ from the first generation:

#1=^A=upper case Greek pi  
 #2=^B=less than  
 #3=^C=greater than  
 #4=^D=a one point space  
 #5=^E=symbol for pound Sterling  
 #6=^F=integral sign  
 #7=^G=contour integral sign  
 #10 through #12 unassigned  
 #13=^K=paragraph symbol  
 #14 through #16 unassigned  
 #17=^O=bullet  
 #20 through #22 unassigned  
 #23=^S=upper case Greek sigma  
 #24 and #25 unassigned  
 #26=^V=three dots meaning because  
 #27 through #37 unassigned  
 #40 through #47 as in first generation  
 #50=prime as in second generation  
 #51=radical symbol, square root symbol  
 #52 through #137 as in first generation  
 #140 unassigned  
 #141 through #176 as in first generation.

### Concluding remarks:

All existing Alto fonts for TimesRoman, Helvetica, and Math have been brought up to date with the third generation special characters, and put on the directory [Ivy]<AltoFonts>; see the AltoFontGuide memo. Clover was running with second generation TimesRoman and Helvetica and first generation Math; the cataclysm has brought all three families up to the third generation.