Date: Oct. 30, 2011
To: Unicode Technical Committee
Subject: Emoji Variation Sequences
From: Peter Edberg, with input from Asmus Freytag, Ned Holbrook, Yasuo Kida, Mark Davis, Kat Momoi, Murray Sargent, and Ken Lunde (though I don't claim that they agree with this document or that it accurately represents their input).

A. Introduction

As part of encoding the emoji core set for Unicode 6.0, 114 emoji from that core set were mapped to (unified with) existing characters or character sequences from Unicode 5.2 or earlier. The associated document "Unified Emoji Reference" provides more detail on the representation across various systems of these 114 emoji.

When these Unicode characters result from text created as emoji, such as e-mail or text messages generated on a Japanese cell phone or via an emoji input palette, users may expect them to be displayed in other contexts using an emoji style similar to the one used when they were created. Such an emoji style usually involves color display, and may sometimes involve other stylistic differences and/or animation.

However, since these Unicode characters were in Unicode prior to the emoji additions — with some dating back to Unicode 1.1 — many systems have already supported such characters using a more traditional black-and-white glyph, and users who created documents with characters displayed this way expect them to remain displayed this way (even when interchanged display).

For example: For U+2764 HEAVY BLACK HEART (in the Dingbats block) the Unicode charts (and Dingbats fonts) use the first of these glyphs while the other glyphs are

used by Japanese cell phones and emoji fonts:

These characters may be sent as part of SMS text messages and e-mail subjects where font information is not available, and a given character may be used with both an emoji style and a non-emoji style in the same document.

Even for the other 654 emoji mapped to new Unicode 6.0 characters, black & white fonts are being created and for some applications there is a need to distinguish between a black & white presentation and an emoji-style presentation. However, for these there is less of a backwards-compatibility display issue.

This document proposes the use of a pair of variation selectors which can be used to specify that the range of glyphs used to represent the character(s) to which they are applied should be restricted either to glyphs with the traditional black-and-white appearance, or to glyphs using an "emoji style", usually involving color and/or possibly animation. The "emoji style" for a given character may include different glyph shapes, different glyph orientations, different glyph colors, etc. There are several possible

implementation options, as described in section B.

In the absence of any such variation selector associated with one of these characters, a display system may choose a glyph based on other context information (perhaps using emoji style in e-mail and text messages, and traditional style in other documents), or may always prefer one style.

Many display systems will not have a color/emoji-style glyph associated with all of these characters and sequences, and may thus display a traditional black & white glyph even when presented with a variation sequence requesting an emoji-style glyph.

B. Implementation options

1. Explicit variation sequences for just the unified emoji

Add to StandardizedVariants.txt an explicit set of variation sequences covering just the Unicode characters used for the 114 unified emoji, with exclusions and considerations as noted in section C. For example:

```
2764 FE00; black & white; # HEAVY BLACK HEART
2764 FE01; color; # HEAVY BLACK HEART
```

A complete list of the characters for which such variation sequences are proposed, together with sample glyphs for each variation selector, is at the end of this document. This is the minimum requirement for this proposal.

2. Explicit variation sequences for all emoji

Add to StandardizedVariants.txt an explicit set of variation sequences covering the Unicode characters used for all 768 emoji in the core set, with exclusions and considerations as noted in section C.

3. Define a new property, apply variation selectors to characters that have it

(Ned, Asmus:) Define a new property designating either "emoji" or something like "dingbats" . (Asmus:) Then instead of having a long list of specific entries in Standardized Variants.txt, allow this pair of variation selectors to be used with all characters that have this property; "An 'emoji style' display should be open to *all* dingbats, because Unicode should not be in the business of tracking which of the dingbats happen to be supported in this way, only whether a character — by its nature — is one that is (also) used as a dingbat.

(Asmus:) "The dingbat property could be enumerated, rather than boolean, so it can be used to separate out characters that are primarily dingbats from characters that are often used as dingbats (e.g. @ or #. or !)."

The set of characters with this property could be the set from 1 or 2 above (for which we already have data in EmojiSources.txt), but would likely be a more generic set that includes those characters.

(Asmus:) "Any messing with variation selectors needs to be very limited. They do

complicate stuff much more than adding some new 'near duplicate' character codes. Having said that, a generic VS to request a 'graphical/color/3D' rendition vs. 'printer's dingbat' rendition seems to make sense.

"I deliberately suggested creating a new property. This behavior of substituting images for glyphs is something that could apply in principle to any dingbat. There's no call to design this mechanism in a way to make it a one-off just for 'emoji'. Once devices supporting Unicode (instead of Shift-JIS) become widespread or once other areas get into the cute icon business, you'll see pressures for the same mechanism."

KIda-san: "I like the idea of having two VSes that have constant meaning across emoji."

4. Reserve two variation selectors as generally applicable for this

Reserve two variation selectors, say U+FE0E and U+FE0F, for general use to designate black & white versus color presentation.

Note: Mark suggested that these should be reserved only for use with symbols. However, several Unicode characters used for unified emoji (and not excluded below) are not symbols, e.g.:

```
2139;INFORMATION SOURCE;Ll;
203C;DOUBLE EXCLAMATION MARK;Po;
2049;EXCLAMATION QUESTION MARK;Po;
```

Note: From *The Unicode Standard,* section 16.4, "The variation selector is not used as a general code extension mechanism; only certain sequences are defined, as follows…".

C. Exclusions and considerations

1. Unified characters that do not need variation sequences/selectors

The following characters are in one or more of the carrier emoji sets primarily as extensions to SJIS rather than as emoji per se. There is no need to have a black&white/color variation sequence pair for these:

  a) Spaces (only in KDDI set):
```
2002;EN SPACE
2003;EM SPACE
2005;FOUR-PER-EM SPACE
```
  b) The symbols ©, ®, and ™ are in all three carrier sets. In two they are presented as normal black & white glyphs. In the KDDI set these are displayed in blue, but since they are mapped to the black & white glyphs for other carriers, it does not seem important to distinguish an emoji style for these:
```
00A9;COPYRIGHT SIGN
00AE;REGISTERED SIGN
2122;TRADE MARK SIGN
```
  c) The punctuation mark `3030;WAVY DASH` is only in the DoCoMo set, and as a normal

black & white punctuation character.

2. Unified emoji mapped to combining sequences; restrictions in TUS 16.4

Eleven of these 114 core emoji, various color versions of ⓪-⑨ and #⃣, are mapped to sequences of Unicode characters (a base character plus a combining enclosing mark), rather than to single characters. The logical variation sequences for these would be, e.g.:

```
0030 20E3 FE00; black & white; # DIGIT ZERO + COMBINING ENCLOSING KEYCAP
0030 20E3 FE01; color emoji; # DIGIT ZERO + COMBINING ENCLOSING KEYCAP
```

However, *The Unicode Standard* section 16.4 says "A variation sequence always consists of a base character followed by a variation selector character" (where "the base character in a variation sequence is never a combining character or a decomposable character."). This seems unnecessarily restrictive; variation selectors should be applicable to sequences. Unless that text is changed, it may not be possible to use variation selectors with these sequences.

Note, an alternative might be to put the variation selector after the base character:

```
0030 FE00 20E3; black & white; # DIGIT ZERO + COMBINING ENCLOSING KEYCAP
0030 FE01 20E3; color emoji; # DIGIT ZERO + COMBINING ENCLOSING KEYCAP
```

But since a variation sequence ends with the variation selector, the actual variation sequence here just applies to the digit, and expresses a distinction that is not relevant for the digit alone. Also, per section 16.4, "The variation selector affects *only* the appearance of the base character."

3. Regional indicator sequences

The 10 flags in the core emoji set are the normal emoji representation for certain pairs of regional indicators from the range U+1F1E6 ..U+1F1FF. Architecturally, variation selectors could not be used to select between emoji and non-emoji representations for those pairs, since they consist of two base characters. If such pairs need to be rendered as separate regional indicators, a ZWNJ can be used in between the two characters.
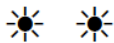
D. Possible further distinctions
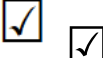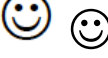
Preliminary discussions for this proposal raised the possibility of using variation selectors to make additional distinctions among different emoji currently mapped to the same Unicode characters, and the possibility of changing the encoding of some emoji. These are not part of this proposal, but material from those discussions is covered in a separate document "Further Emoji Distinctions."

E. Variation sequences for unified characters

The following table gives Unicode characters (and sequences) with the proposed corresponding variation selectors and sample glyphs for those variation selectors, per the "minimal" implementation option in section B.1 above. The variations could be

distinguished as traditional style vs. emoji style, text style vs. graphical icon style, or (as used below) black & white vs. color. For the black & white variations, the first sample glyph is the one from the Unicode 6.0 standard; the others (if any) are from Apple fonts including Apple Symbols, Zapf Dingbats, and the Hiragino family.

Unicode character/sequence	Variation selectors	Sample glyphs
2139;INFORMATION SOURCE	FE00; black & white;	i i i
	FE00; color;	i
203C;DOUBLE EXCLAMATION MARK	FE00; black & white;	‼ ‼ ‼
	FE00; color;	‼ ‼
2049;EXCLAMATION QUESTION MARK	FE00; black & white;	⁉ ⁉ ⁉
	FE00; color;	⁉ ⁉
231A;WATCH	FE00; black & white;	⌚ ⌚ ⌚
	FE00; color;	⌚ ⌚
231B;HOURGLASS	FE00; black & white;	⌛ ⌛ ⌛
	FE00; color;	⌛
24C2;CIRCLED LATIN CAPITAL LETTER M	FE00; black & white;	Ⓜ Ⓜ
	FE00; color;	Ⓜ
2600;BLACK SUN WITH RAYS	FE00; black & white;	☀ ☀
	FE00; color;	☀ ☀ ☀ ☀
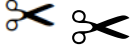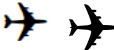
2601;CLOUD	FE00; black & white;	
	FE00; color;	
260E;BLACK TELEPHONE	FE00; black & white;	
	FE00; color;	
2611;BALLOT BOX WITH CHECK	FE00; black & white;	
	FE00; color;	
2614;UMBRELLA WITH RAIN DROPS	FE00; black & white;	
	FE00; color;	
2615;HOT BEVERAGE	FE00; black & white;	
	FE00; color;	
261D;WHITE UP POINTING INDEX	FE00; black & white;	
	FE00; color;	
263A;WHITE SMILING FACE	FE00; black & white;	
	FE00; color;	

2648;ARIES = ram	FE00; black & white;	♈ ♈
	FE00; color;	♈ 🐑 ♈ ♈
2649;TAURUS = bull	FE00; black & white;	♉ ♉
	FE00; color;	♉ 🐂 ♉ ♉
264A;GEMINI = twins	FE00; black & white;	♊ ♊
	FE00; color;	♊ 🦋 ♊ ♊
264B;CANCER = crab	FE00; black & white;	♋ ♋
	FE00; color;	♋ 🦀 ♋ ♋
264C;LEO = lion	FE00; black & white;	♌ ♌
	FE00; color;	♌ 🦁 ♌ ♌
264D;VIRGO = virgin maiden	FE00; black & white;	♍ ♍
	FE00; color;	♍ 💁 ♍ ♍
264E;LIBRA = scales	FE00; black & white;	♎ ♎
	FE00; color;	♎ ⚖ ♎ ♎

264F;SCORPIUS = scorpion	FE00; black & white;	♏ ♏
	FE00; color;	♏🦂♏♏
2650;SAGITTARIUS = archer	FE00; black & white;	♐ ♐
	FE00; color;	♐🏹♐♐
2651;CAPRICORN = goat	FE00; black & white;	♑ ♑
	FE00; color;	♑🐐♑♑
2652;AQUARIUS = water carrier	FE00; black & white;	♒ ♒
	FE00; color;	♒🏺♒♒
2653;PISCES = fish	FE00; black & white;	♓ ♓
	FE00; color;	♓🐟♓♓
2660;BLACK SPADE SUIT	FE00; black & white;	♠ ♠ ♠
	FE00; color;	♠♠♠♠
2663;BLACK CLUB SUIT	FE00; black & white;	♣ ♣ ♣
	FE00; color;	♣♣♣♣
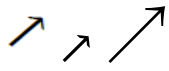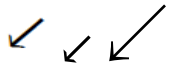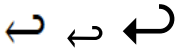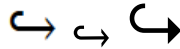
2665;BLACK HEART SUIT	FE00; black & white;	♥ ♥ ♥
	FE00; color;	♥ ♥ ♥ ♥
2666;BLACK DIAMOND SUIT	FE00; black & white;	♦ ♦ ♦
	FE00; color;	♦ ♦ ♦ ♦
2668;HOT SPRINGS	FE00; black & white;	♨ ♨ ♨
	FE00; color;	♨ ♨ ♨ ♨
267B;BLACK UNIVERSAL RECYCLING SYMBOL	FE00; black & white;	♻ ♻ ♻
	FE00; color;	♻ ♻
267F;WHEELCHAIR SYMBOL	FE00; black & white;	♿ ♿
	FE00; color;	♿ ♿ ♿ ♿
2693;ANCHOR	FE00; black & white;	⚓ ⚓
	FE00; color;	⚓
26A0;WARNING SIGN	FE00; black & white;	⚠ ⚠
	FE00; color;	⚠ ⚠ ⚠ ⚠
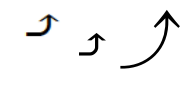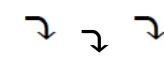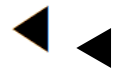
26A1;HIGH VOLTAGE SIGN	FE00; black & white;	
	FE00; color;	
26BD;SOCCER BALL	FE00; black & white;	
	FE00; color;	
26BE;BASEBALL	FE00; black & white;	
	FE00; color;	
26C4;SNOWMAN WITHOUT SNOW	FE00; black & white;	
	FE00; color;	
26C5;SUN BEHIND CLOUD	FE00; black & white;	
	FE00; color;	
26D4;NO ENTRY	FE00; black & white;	
	FE00; color;	
26EA;CHURCH	FE00; black & white;	
	FE00; color;	
26F2;FOUNTAIN	FE00; black & white;	
	FE00; color;	
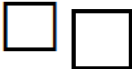
26F3;FLAG IN HOLE =golf	FE00; black & white;	
	FE00; color;	
26F5;SAILBOAT	FE00; black & white;	
	FE00; color;	
26FA;TENT	FE00; black & white;	
	FE00; color;	
26FD;FUEL PUMP = gas station	FE00; black & white;	
	FE00; color;	
2702;BLACK SCISSORS	FE00; black & white;	
	FE00; color;	
2708;AIRPLANE	FE00; black & white;	
	FE00; color;	
2709;ENVELOPE	FE00; black & white;	
	FE00; color;	
270C;VICTORY HAND	FE00; black & white;	
	FE00; color;	

270F;PENCIL	FE00; black & white;	✏️ ✏️
	FE00; color;	✏️ ✏️
2712;BLACK NIB	FE00; black & white;	✒️ ✒️
	FE00; color;	✒️ ✒️
2714;HEAVY CHECK MARK	FE00; black & white;	✔ ✔
	FE00; color;	✔️
2716;HEAVY MULTIPLICATION X	FE00; black & white;	✖ ✖
	FE00; color;	✖️
2733;EIGHT SPOKED ASTERISK	FE00; black & white;	✳ ✳
	FE00; color;	✳️ 🔲 ✴️
2734;EIGHT POINTED BLACK STAR	FE00; black & white;	✴ ✴
	FE00; color;	✦ ✴️ ✴️
2744;SNOWFLAKE	FE00; black & white;	❄ ❄
	FE00; color;	❄️
2747;SPARKLE	FE00; black & white;	❇ ❇
	FE00; color;	❇️ (animated)

2757;HEAVY EXCLAMATION MARK SYMBOL	FE00; black & white;	!
	FE00; color;	! ! ! !
2764;HEAVY BLACK HEART	FE00; black & white;	❤ ❤
	FE00; color;	❤ 💗 💗 ❤
27A1;BLACK RIGHTWARDS ARROW	FE00; black & white;	➡ ➡ ➡
	FE00; color;	➡ ➔ ➡
2B05;LEFTWARDS BLACK ARROW	FE00; black & white;	← ← ←
	FE00; color;	⬅ ← ⬅
2B06;UPWARDS BLACK ARROW	FE00; black & white;	↑ ↑ ⬆
	FE00; color;	⬆ ↑ ⬆
2B07;DOWNWARDS BLACK ARROW	FE00; black & white;	↓ ↓ ⬇
	FE00; color;	⬇ ↓ ⬇
2196;NORTH WEST ARROW	FE00; black & white;	↖ ↖ ↖
	FE00; color;	↖ ↖ ↖ ⬉

2197;NORTH EAST ARROW	FE00; black & white;	↗ ↗ ↗
	FE00; color;	↗ ↗ ↗ ↗
2198;SOUTH EAST ARROW	FE00; black & white;	↘ ↘ ↘
	FE00; color;	↘ ↘ ↘ ↘
2199;SOUTH WEST ARROW	FE00; black & white;	↙ ↙ ↙
	FE00; color;	↙ ↙ ↙ ↙
2194;LEFT RIGHT ARROW	FE00; black & white;	↔ ↔ ↔
	FE00; color;	↔ ↔
2195;UP DOWN ARROW	FE00; black & white;	↕ ↕ ↕
	FE00; color;	↕ ↕
21A9;LEFTWARDS ARROW WITH HOOK	FE00; black & white;	↩ ↩ ↩
	FE00; color;	↩ ↩
21AA;RIGHTWARDS ARROW WITH HOOK	FE00; black & white;	↪ ↪ ↪
	FE00; color;	↪

2934;ARROW POINTING RIGHTWARDS THEN CURVING UPWARDS	FE00; black & white;	⤴ ⤴ ⤴
	FE00; color;	⤴ ⤴
2935;ARROW POINTING RIGHTWARDS THEN CURVING DOWNWARDS	FE00; black & white;	⤵ ⤵ ⤵
	FE00; color;	⤵ ⤵
25AA;BLACK SMALL SQUARE	FE00; black & white;	▪ ▪ ▪
	FE00; color;	◼
25AB;WHITE SMALL SQUARE	FE00; black & white;	▫ ▫ ▫
	FE00; color;	◻
25B6;BLACK RIGHT-POINTING TRIANGLE	FE00; black & white;	▶ ▶
	FE00; color;	▶ ▷ ▶
25C0;BLACK LEFT-POINTING TRIANGLE	FE00; black & white;	◀ ◀
	FE00; color;	◀ ◁ ◀
25FB;WHITE MEDIUM SQUARE	FE00; black & white;	◻ ◻ ◻
	FE00; color;	◼
25FC;BLACK MEDIUM SQUARE	FE00; black & white;	◼ ◼ ◼
	FE00; color;	◼

25FD;WHITE MEDIUM SMALL SQUARE	FE00; black & white;	□ □ ☐
	FE00; color;	🟧
25FE;BLACK MEDIUM SMALL SQUARE	FE00; black & white;	▪ ▪ ■
	FE00; color;	🟩
26AA;MEDIUM WHITE CIRCLE	FE00; black & white;	○ ○ ○
	FE00; color;	🔴
26AB;MEDIUM BLACK CIRCLE	FE00; black & white;	● ● ●
	FE00; color;	🔵
2B1B;BLACK LARGE SQUARE	FE00; black & white;	■ ■
	FE00; color;	🟩
2B1C;WHITE LARGE SQUARE	FE00; black & white;	☐ ☐
	FE00; color;	🟧
2B50;WHITE MEDIUM STAR	FE00; black & white;	☆ ☆
	FE00; color;	⭐ 🌟 ⭐

2B55;HEAVY LARGE CIRCLE	FE00; black & white;	○
	FE00; color;	○ ○ ○
303D;PART ALTERNATION MARK	FE00; black & white;	〽 〽
	FE00; color;	〽 〽
3297;CIRCLED IDEOGRAPH CONGRATULATION	FE00; black & white;	㊗ ㊗
	FE00; color;	㊗ ㊗ ㊗
3299;CIRCLED IDEOGRAPH SECRET	FE00; black & white;	㊙ ㊙
	FE00; color;	㊙ ㊙ ㊙ ㊙
1F004;MAHJONG TILE RED DRAGON	FE00; black & white;	🀄
	FE00; color;	🀄 🀄 🀄
1F17F;NEGATIVE SQUARED LATIN CAPITAL LETTER P	FE00; black & white;	🅿 🅿
	FE00; color;	🅿 🅿 🅿 🅿
1F21A;SQUARED CJK UNIFIED IDEOGRAPH-7121	FE00; black & white;	🈚
	FE00; color;	🈚 🈚

1F22F;SQUARED CJK UNIFIED IDEOGRAPH-6307	FE00; black & white;	指
	FE00; color;	指 指 指

Unicode sequences, may not be able to use variation selectors with these (and note, the Unicode charts charts do not show glyphs for these sequences).

0023 20E3;NUMBER SIGN + COMBINING ENCLOSING KEYCAP	FE00; black & white;	#
	FE00; color;	# # # #
0030 20E3;DIGIT ZERO + COMBINING ENCLOSING KEYCAP	FE00; black & white;	0
	FE00; color;	0 0 0 0
0031 20E3;DIGIT ONE + COMBINING ENCLOSING KEYCAP	FE00; black & white;	1
	FE00; color;	1 1 1 1
...		
0038 20E3;DIGIT EIGHT + COMBINING ENCLOSING KEYCAP	FE00; black & white;	8
	FE00; color;	8 8 8 8
0039 20E3;DIGIT NINE + COMBINING ENCLOSING KEYCAP	FE00; black & white;	9
	FE00; color;	9 9 9 9