

Proposed enhancements for emoji characters: background

To: UTC
Date: 2014 August 03
From: Peter Edberg, Mark Davis
Live Doc <http://goo.gl/g3lBiQ>

Emoji characters have become extremely popular. Yet the choice of emoji to be represented in Unicode has left many people confused or disappointed. People also misunderstand how long it takes to encode characters (and support them on operating systems), and what terms like WHITE in character names mean. (For more on this see the Unicode FAQ.) Yet it is clear that the Unicode Consortium needs to address more quickly some of the issues that have come up.

This proposal attempts to address several issues related to emoji-like characters in Unicode:

- Concerns about the lack of diversity in the representation of the people and body part emoji.
- Concerns about apparent bias in the selection of encoded characters, and/or notable gaps in the current set.
- Popular requests for emoji additions.

This document is one of a set which should be read and considered together:

- [L2/14-172](#) Proposed enhancements for emoji characters: background
- [L2/14-173](#) Variation selectors for Emoji skin tone
- [L2/14-174](#) Emoji Additions

It is also recommended that people read the working draft *Unicode Emoji* at unicode.org/reports/tr51/ before continuing.

Contents

[A. Sources of input](#)

[B. Long-term direction: embedded graphics](#)

[C. People & body part emoji, default appearance and diversity](#)

[C1. Default presentation for generic people & body part emoji](#)

[C2. Supporting and interchanging multiple presentations for existing people & body part emoji](#)

[C3. Challenge: Multi-person groupings and variation selector](#)

[D. Additional characters](#)

A. Sources of input

- Concerns and requests expressed directly to the Unicode Consortium and to member organizations
- Popular petitions and campaigns related to emoji, notably:
 - Campaigns for more diversity (see details in diversity section below).
 - Campaigns for specific food items: taco and hot dog (see details in food section below).
- Websites with collections of requests for new emoji (some surprising consistency among these and other requests), e.g.:
 - [Life Won't Be Complete Until We Get These Emojis](#) (NYmag)
 - [19 Emoji That Really Should Exist](#) (Business Insider)
 - [18 Emojis That Should Exist But Don't](#) (Buzzfeed)

B. Long-term direction: embedded graphics

Unicode will never be able to encode all of the wide variety of emoji-like images that people seem to want to use in message, documents, and social media. The longer-term solution is to facilitate including these as embedded graphics (aka “stickers”). However:

- There are protocol and other restrictions that limit where these can be used (e.g. not currently in e-mail subject lines, SMS messages, or filenames (images can be included in a Twitter message via a URL that references the image on an external server).
- The user interface for entering images is typically much more cumbersome than an emoji keyboard or palette.
- Such images do not necessarily scale automatically to match adjacent text.

Platform vendors and protocol developers need to work on improving those issues. However, Unicode should address the highest-priority concerns using encoding-based approaches, not only to provide a quicker solution, but to avoid the appearance of preferred treatment for certain characters within some potentially larger sets.

For more about this, see the working draft *Unicode Emoji* at unicode.org/reports/tr51/

C. People & body part emoji, default appearance and diversity

The physical appearance of people may differ in skin tone, hair style & color, facial hair, eye color & shape, use of prescription glasses, body size and shape, etc. Presenting the generic people and body part emoji such as WOMAN or NOSE with a single set of specific features appears to privilege those particular features; presenting emoji for occupations (CONSTRUCTION WORKER, INFORMATION DESK PERSON) with a particular gender may

be seen as reflecting traditional biases. (Note, here “body part” includes the hand symbols).

Furthermore, people do seem to want to use and reliably interchange emoji that reflect at least some of the appearance variation present in real people.

Some related links:

- [Apple and Google: Support Equality. Make Diverse Emojis!](#)
- [Apple: Add More Diversity to the Emoji Keyboard](#)
- [Unicode Unveils 250 New Emojis, Gets Thumbs Down For Diversity](#)
- [iDiversicons](#)

Let us consider each of the related issues in turn.

C1. Default presentation for generic people & body part emoji

As much as possible, the generic people and body part emoji should be drawn by default—in Unicode charts and in vendor fonts—to avoid identification with specific physical features other than those that are part of the character name (as in PERSON WITH BLOND HAIR)—perhaps closer to outlines or silhouettes, or to the emoticons. The charts (and the Android and Windows fonts) already do this for, e.g., FAMILY, MAN AND WOMAN HOLDING HANDS, etc. When these are shown in color, they should preferably use a non-realistic skin tone (such as yellow-orange). The emoji should be drawn as gender-neutral if the character name does not designate gender (eg, COUPLE and KISS).

→ *Proposal*: Consensus on a goal to change glyphs in the chart fonts as necessary.

C2. Supporting and interchanging multiple presentations for existing people & body part emoji

On the other hand, people want to be able to use and interchange emoji with more specific physical appearance; merely making the generic emoji look generic (i.e. like no real person) does not provide any diversity of appearance, it just avoids privileging certain physical characteristics. An encoding-based approach is not suitable for conveying all of the variations that people might want to represent, and supporting too many specific-appearance variations of each generic emoji would have a huge negative impact on both font size and complexity of input. A reasonable compromise would be to support 3-5 more-specific appearance variations for each generic emoji (systems that do not support these would show the generic appearance). Of course, any move toward encoding some physical appearance variants can raise issues about why other variants are not being encoded, but the argument about font size and input complexity is a compelling response.

→ *Proposal*: Use 3 Variation Selectors to indicate the preferred skin tone among 3 choices. For details, see [L2/14-173](#) *Variation selectors for Emoji skin tone*.

C3. Challenge: Multi-person groupings and variation selector

There are several emoji for multi-person groups:

- KISS (two people kissing)
- COUPLE WITH HEART
- FAMILY (two parents and child[ren])
- MAN AND WOMAN HOLDING HANDS
- TWO MEN HOLDING HANDS
- TWO WOMEN HOLDING HANDS

To better represent the variety of real couples and families (including adopted children), it would be desirable not only to use the above variation selectors to affect all members of the grouping, but to at least be able to specify light or dark skin tone separately for different members of the grouping.

→ *Proposal*: Use 2 Variation Selectors to pick indicate some additional combinations of tones. For details, see [L2/14-173 Variation selectors for Emoji skin tone](#).

D. Additional characters

As emoji have become more popular, the selection process for new emoji (aka “pictographs that could have an emoji representation”) has puzzled and disappointed people. Up until now, most of the emoji characters have been encoded “reactively”, based on existing sets. To some extent, however, we have also encoded characters that “flesh out a set” rather than simply incorporating characters in existing sets: see unicode.org/reports/tr51/#Encoding .

The [L2/14-174 Emoji Additions](#) document proposes new characters in various categories. The focus is on addressing perceptions of bias, filling in notable gaps, and responding to highly popular requests. As always, adding some characters can prompt questions about why other potential candidates were not added. There are several factors that went into the selection of these proposed characters.

1. Does the proposed pictograph fill in a perceived gap in existing types of emoji? Does the proposed pictograph “fill in a hole” among that type of character? For example, we have TIGER, but not LION.
2. Do we appear right now to be “playing favorites”? For example, Unicode has a CHURCH but no MOSQUE.
3. Is there a clearly recognizable image that could serve as a paradigm, standing for a range of possible images?
4. Is the expected frequency of use at least as high as current emoji characters of the same type? We’ve looked at various measures of this, such as frequency of associated words in web pages.
5. Is it often requested (eg HOT DOG or UNICORN)?
6. Is it just one of many types, with no special reason to favor it? For example, there are thousands of occupations: is there a special reason to favor particular ones of them? (DOCTOR, DENTIST, JANITOR, etc.)

For rapid adoption, these characters may initially be assigned in a private-use-area registry, but the further goal is that they would eventually be standardized. The PUA characters could initially be described in a section of TR51, allowing them to also have associated default values and

annotations.

The number of added characters in a first set have been winnowed down from about 200 to about 50, to address concerns about font size and input complexity. This is an initial proposal, and we can refine it further based on feedback.

The specific proposed additions are discussed in a separate document: [L2/14-174](#) *Emoji Additions*.