

Gender Emoji ZWJ Sequences

M. Davis, R. Been, and the Unicode Emoji Subcommittee

2016-07-14

Working draft: goo.gl/yRXj3S

This document describes how vendors can support a set of both female and male versions of many emoji characters, including new profession emoji. Because these emoji use sequences of existing Unicode characters composed according to [UTR#51: Unicode Emoji](#), vendors can begin design and implementation work now and can deploy before the end of 2016, rather than waiting for Unicode v10.0 to come out in June of 2017.

The set of gender emoji in this document has been approved by the Unicode Emoji Subcommittee, and is designed for interoperability across vendors.

Contents

[Professions](#)

[Gender Alternates](#)

[Skin Tone Modifiers](#)

[Emoji Variation Selectors](#)

[Listed Sequences](#)

[Counts](#)

[Background](#)

[Final 11 Professions](#)

[Design Considerations](#)

Introduction

Unicode itself does not normally specify the gender for emoji characters: the emoji character is RUNNER, not MAN RUNNING; POLICE OFFICER not POLICEMAN. Even where the name may appear to be exclusively one gender, such as U+2603 SNOWMAN or U+1F482 GUARDSMAN the character can be treated as neutral regarding gender.

To get a greater sense of realism for these characters, however, vendors typically have picked the appearance of a particular gender to display. This has led to gender disparities in the emoji that people can use. There is also a lack of emoji representing professions and roles, and the few that are present (like POLICE OFFICER) do not provide for both genders; a vendor has to choose one or the other, but can't represent both.

Thus U+1F575 SLEUTH OR SPY (detective) is typically displayed as male, while the U+1F64D PERSON FROWNING is typically displayed as female. Vendors cannot have *both* a male and female version of each one.



1F575



1F64D

The Unicode Technical Committee (UTC) and the Unicode Emoji Subcommittee (ESC) have been working to address gender issues for some time, resulting in the additional characters in Unicode 9.0 (June 2016) to complete gender

pairs, such as PRINCE (for PRINCESS), and some additional roles such as PREGNANT WOMAN. But the process of adding new characters is slow (these characters were the result of [work done in 2014](#)), and so the UTC investigated more generative mechanisms, such as in [TR#52](#). However, it had become clear that the TAG approach in the draft [TR#52](#) had some significant disadvantages, notably for fallback on older systems. That is, it will be some time before the fallback image for unsupported TAG sequences would be supported, and in the meantime people expecting to interchange a female version of an emoji would have recipients get a male version (or vice versa).

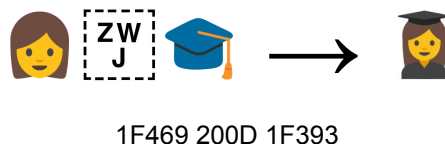
During UTC meeting #147, we decided that a better solution to gender than TAGs would be *emoji zwj sequences* (ZWJ = [Zero-Width Joiner](#)). Essentially, each emoji zwj sequence is treated like a ligature, resulting in a single glyph being displayed on the screen. Keyboard input, segmentation, and other operations are modified to handle the sequence as a whole unit. These emoji zwj sequences are already used to represent diverse family groupings and others, and could be extended relatively quickly.

The choice of which emoji zwj sequences to support is up to vendors, but the Unicode Emoji Subcommittee was tasked with getting consensus on a set that would meet the immediate needs of vendors in order to promote interoperability. The ESC has included both a set of professions using emoji sequences, based on the [Google proposal](#), and a set of gender alternates for current characters, based on work done in [TR#52](#).



Note that the images used below are simply for illustration: vendors will produce images to fit with their overall design.

Professions

Professions (roles) are represented by using a MAN or WOMAN as the initial character, followed by a ZWJ, followed by a representative object character, such as in the following for a *student*.



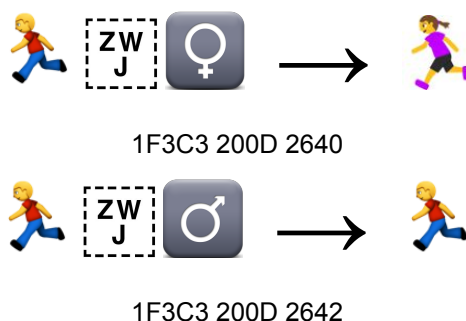
One existing character (the [symbol for health care](#)) will need to have its Emoji property value set to Yes in the data files for TR #51 v4.0, and be added to StandardizedVariants in Unicode 10.0.

		U+2695	STAFF OF AESCULAPIUS
---	---	------------------------	----------------------





As above, each sequence should be treated as a whole in keyboards and palettes, though vendors might use mechanisms to reduce the “scroll” in keyboard palettes, such as by using *long-press* as currently done for skin tone modifier sequences. Vendors need not add this additional character to their emoji palettes/keyboards.

Gender Alternates

Gender alternates are represented by an initial character for a person, plus ZWJ, plus the FEMALE SIGN or MALE SIGN. Thus an explicitly gendered version of RUNNER is specified by the following two sequences:



Two existing characters will need to have their Emoji property values set to Yes in the data files for TR #51 v4.0, and be added to StandardizedVariants.txt in Unicode 10.0.

		U+2640	FEMALE SIGN
		U+2642	MALE SIGN

Each sequence has an explicit gender determined by one of these two signs, independent of what the default gender for the initial character (eg, RUNNER) would be on the platform. Vendors are encouraged to use as neutral a representation as possible for the unadorned initial character. However, current designers have found that difficult to represent. The current focus is primarily to provide improved representation of women (and men) by allowing the vendors to use both male and female versions of emoji characters, without being forced to choose only one.

Some emoji are commonly depicted with two people, such as U+1F46F WOMAN WITH BUNNY EARS (aka *women partying*) and 1F93C WRESTLERS. For such characters, the gender zwj sequences would show the people as either

1. both male (with U+2642 MALE SIGN), or
2. both female (with U+2640 FEMALE SIGN).

Each sequence should be treated as a whole in keyboards and palettes, though vendors might use mechanisms to reduce the “scroll” in keyboard palettes, such as by using *long-press* as currently done for skin tone modifier sequences. Vendors need not add the two additional signs to their emoji palettes/keyboards.

Skin Tone Modifiers

Vendors can also support skin tones, by adding emoji modifier characters after the initial character. For a Fitzpatrick type 5 female student and a type 1-2 female runner, the following sequences would be used.



1F469 1F3FF 200D 1F393



1F3C3 1F3FB 200D 2640

Emoji Variation Selectors

Any character that has the property value [Emoji Presentation=No](#) should be followed by an Emoji Variation Selector (VS16 = U+FE0F) in emoji zwj sequences. This is also the form strongly recommended for keyboards, for good fallback behavior on older systems. *The one exception is that the VS16 must be omitted before and after a skin modifier.*

Correct



26F9 FE0F 200D 2640 FE0F



26F9 1F3FF 200D 2640 FE0F

Incorrect



26F9 1F3FF FE0F 200D 2640 FE0F



26F9 FE0F 1F3FF 200D 2640 FE0F

Display

Fonts, segmentation, and other processing should support emoji zwj sequences both with *and* without the VS16 for more robust interchange. For example, both of the following should work in fonts:



1F3C3 200D 2640



1F3C3 200D 2640 FE0F

Listed Sequences

The proposed sequences are listed in [16182-gender-zwj-sequences-list.pdf](#), with the following columns. The characters will be replaced by images once available, and then added to the [chart of cataloged emoji ZWJ sequences](#).

Key

Code Sequence	The full code sequence, with VS16 codes if applicable
Without VS16	The alternate sequence without VS16 that should be supported in display (if there is one)
Characters	The list of characters
Short name	A short name for CLDR annotations
Annotations	Keywords for CLDR annotations
Image file	The recommended file name for image files

The spreadsheet includes one gender-related zwj sequence that is discussed in a separate proposal: the rainbow flag. That is presented in a [separate document](#). The list includes the skin-tone modifiers in the lists of file names on the second sheet, although those can be derived algorithmically using the Unicode emoji properties.

To see some of the images used by current vendors for the initial characters, see [full-emoji-list](#).

Counts

The [character counts in TR51](#) will be updated by adding 3 to the Singleton count, plus the following new rows. There are a number of new ZWJ sequences that typically have the same image as some singleton or modifier sequence. For example, the explicitly male Sleuth could look the same as the default Sleuth. The count of these is listed under “Typically duplicating sequence” below.

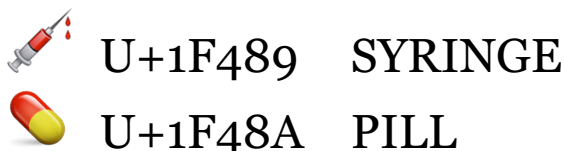
Thus even though the total number of cataloged ZWJ sequences would grow by 508, typically the number of *distinct* new images would be 320.

Type	Count	Subtotal
Emoji ZWJ Sequence: Gendered Role, with object	132	132
Emoji ZWJ Sequence: Gendered Role, with f/m sign	72	204
Emoji ZWJ Sequence: Gendered Activity, with f/m sign	196	400
Emoji ZWJ Sequence: Gendered Gestures, with f/m sign	108	508
Typically duplicating sequence	-188	320

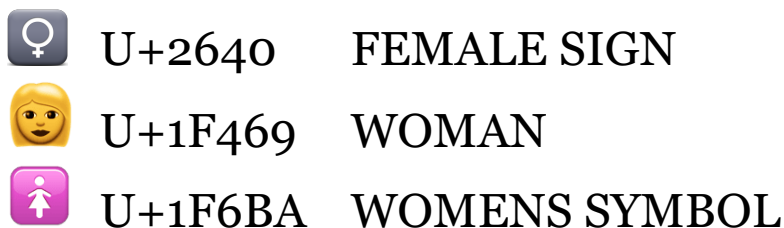
Background

The list is the result of research on the use of current emoji and top professions. This document won't recap those discussions, except that it is worth calling out a few cases, such as why the three characters listed above will have the Emoji property added (for use in these emoji zwj sequences).

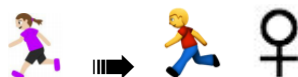
For health professional, there are limited possibilities. Both of the following were considered, but were viewed as not being particularly representative of people involved in health care.



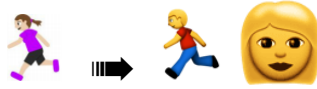
We considered 3 candidates for the male and female symbols to use in the ZWJ sequences:



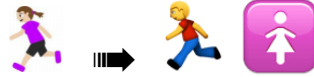
We chose the U+2640 FEMALE SIGN, which has the advantage of being very abstract, and thus the fallback representation on older devices would be very unlikely to be confused with a sequence of two separate entities. The fallback presentation would be something like the following.



The U+1F469 WOMAN character has the disadvantage is that the fallback could be more easily confused with the user's meaning two separate people, such as *a man running from a woman*.



The U+1F6BA WOMENS SYMBOL is more abstract, and so less likely to be taken as a separate entity. However, it is also used as a symbol for restroom, so the fallback could be taken as the fallback could be more easily confused with the user's meaning two separate entities, such as *a man running from a women's restroom*.



For gender alternates, we started with the emoji modifier bases (characters that could take skin-tone modifiers). Characters that already had gender pairs such as U+1F466 BOY and U+1F467 GIRL were removed. Characters without strong gender distinctions such as U+1F476 BABY and U+270C VICTORY HAND were removed, because any differences are too small to show at small sizes. Two characters were added to that list: U+1F3CC GOLFER and U+1F46F WOMAN WITH BUNNY EARS (aka *women partying*), and two specialized characters removed: U+1F930 PREGNANT WOMAN and U+1F472 MAN WITH GUA PI MAO. Note that the specialized character U+1F574 MAN IN BUSINESS SUIT LEVITATING is also not included.

Final 11 Professions

[The original Google proposal](#) proposed 13 representations of women in the workforce. After ensuring the final proposal had wide industry coverage and global applicability, Google and the Unicode Emoji Subcommittee consolidated the final list to 11, deciding not to pursue an additional healthcare representation and a high-tech worker.

We determined that the healthcare category could be represented by a single emoji promoting a myriad of professions within healthcare. For the high-tech worker we determined that at a small size comprehension was difficult. Furthermore there was no ZWJ fallback sequence that adequately translated the meaning. We also modified the appearances of the emoji based on the design considerations below.

As with all emoji, one particular image can be representative of a wide range of uses, such as illustrated by the keywords below.

Thus the final set of professional emoji in this set is as follows. For sample images, see [Design Considerations](#).

Category	Image	Keywords
AGRICULTURE	<i>Farmer</i>	farm worker, rancher, gardener, farming, ...
MANUFACTURING /INDUSTRY	<i>Welder</i>	factory worker, industrial worker, assembly worker, manufacturing, ...
	<i>Mechanic</i>	tradesperson, plumber, electrician, repair person, ...
STEM (Science Technology Engineering & Math)	<i>Health Worker</i>	healthcare professional, doctor, nurse, therapist, ...
	<i>Scientist</i>	Technologist, engineer, mathematician, chemist, physicist, biologist, ...
	<i>Coder</i>	Software developer, engineer, inventor, gamer, ...
SERVICES /EDUCATION	<i>Business Worker</i>	Office worker, business person, financial analyst, manager, architect, white-collar worker, ...
	<i>Chef</i>	Cook, cooking, ...
	<i>Student</i>	Pupil, graduate, ...
	<i>Teacher</i>	Professor, instructor, ...
AWESOMENESS	<i>Rockstar</i>	Singer, entertainer, ...

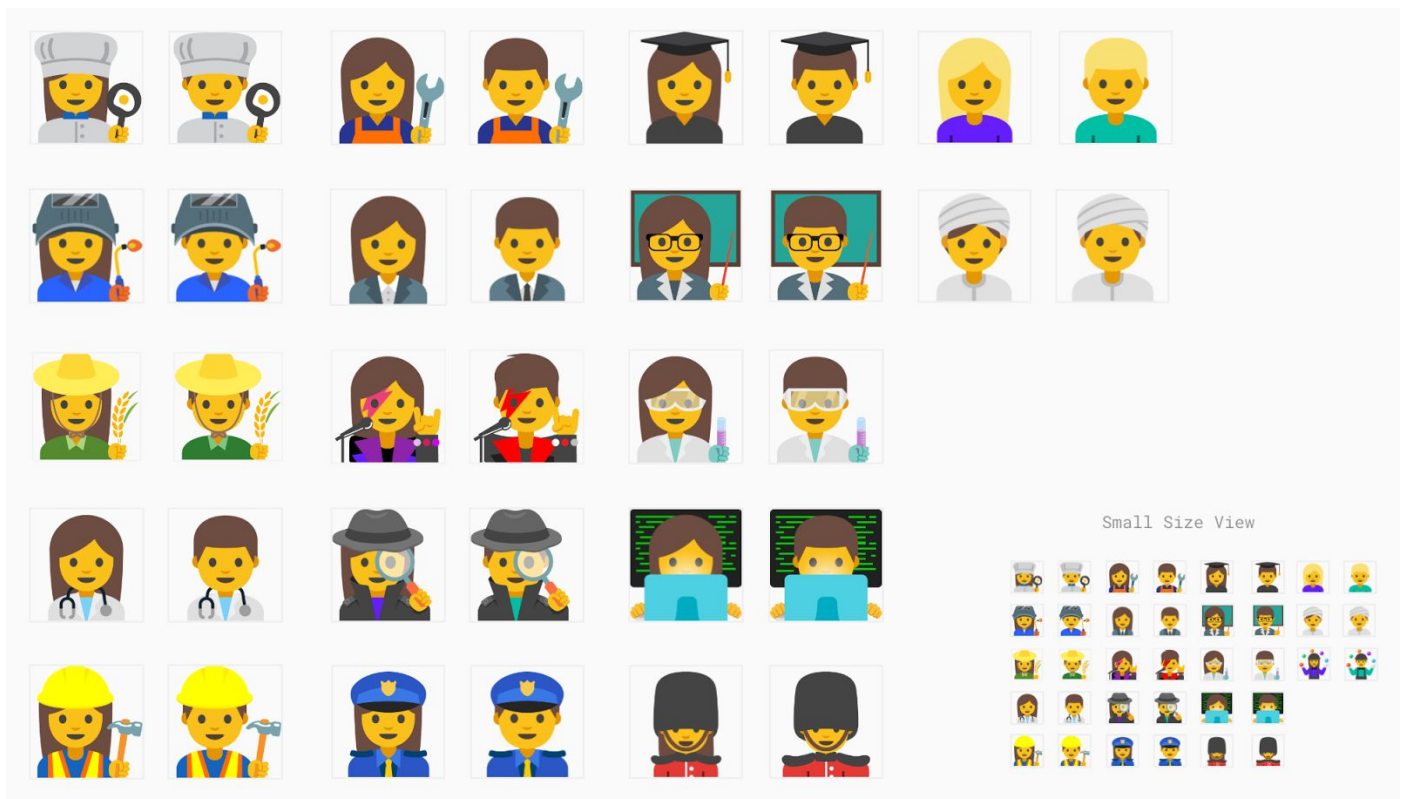
Design Considerations

The following describes some of the design considerations that went into producing a consistent set of images at Google and other vendors. These considerations are not normative, but provide a useful comparison for others' designs. The samples include not only the professions, but some of the gender pairs for existing characters.

The Unicode Emoji Subcommittee and Google's internal global research teams determined readability and appropriateness of the suggested representations in terms of professions, clothing, color, and props. The committee and research teams expressed understanding and positivity towards the final selection. There were initial concerns around considerations such as women's hair appearing uncontained in professions that have safety and health requirements.

Based on these concerns, Google determined that hair was a key representation in attributing gender, and despite the reality of hair standards in certain industries, there would be a consistent style to signal gender. Clothing and colors were selected that were globally comprehensible despite being culturally specific. The selected props were related to the category and as globally representative as possible.

The samples include not only the professions, but some of the gender pairs for existing characters.



Less is More

Google capped the range of representative attributes to **three**: hat or background element, clothing, item/hand posture. The least number of objects needed to represent the profession was ideal.

For example, there was an initial draft of a briefcase in the business woman/man emoji, but the clothing was deemed determinate enough to relay the idea of a “working professional”. Additionally, not including this element allowed for more versatile use.

Background elements had the potential to overcomplicate the emoji. Therefore, the chalkboard and coding screens were simplified to basic shapes with minimal detailing at a close proximity to help signify the profession.

Shoulders

Google and Microsoft looked at head and full body characterizations. Head-only could only signal the profession or role with a hat, which was too limiting for most professions. The full body was unnecessary since a majority of the characterization was associated with the torso, and objects that represented the professions were imperceptible in the full body representations. To maintain consistency, it is cleaner for all professional representations (such as POLICE OFFICER) be converted to shoulder emojis.

Uniformity between Genders

Google minimized the visual delta between genders. In most cases, hair length and tweaks to facial features were the only changes. Per color, culturally biased “female” hues were purposely not selected. The rockstar and “women/men partying” color is different for stylistic reasons.

Object Consistencies

As much as possible, Google maintained scale and object placement consistencies. The hand postures and scale are the same throughout. Objects are similar in size and generally held slightly angled in the left hand.