

Source:	CheonHyeong Sim(沈天珩)
Title:	Proposal to add back U+2E9A
Date:	2022-10-01
Action:	To be considered by UTC (CJK & Unihan Group(?))

In the block CJK Radicals Supplement, there is a hole at the codepoint U+2E9A. The hole is often misunderstood as a removal by many people, but actually it has never been defined since the first occurrence of this block in Unicode.

The misunderstanding mainly comes from that all the other CJK blocks do not have any holes. And in practice, any holes in the CJK blocks are not expected. In the early years, the draft of the compatibility characters from DPRK had a lot of holes because the original characters on those codepoints were unified to the characters submitted by T-source in the compatibility supplement block, but in the final officially released version of Unicode, the characters were reordered and there were no holes anymore; Several months ago, it was proposed that there are two duplicated characters in the CJK Extension H block, since there was no time to reorder the thousands of characters at that period, some people suggested to delete the two characters on their codepoint and leave two holes in the block, but at last the suggestion was also unaccepted.

A widely circulated saying says that this character was a duplication of a character in the Kangxi Radical block. To end this problem, I found the original proposal N1923 (<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n1923.pdf>) to see what on earth happened that time. We could see that the original character on U+2E9A in the proposal is “无”, and considered as a duplication of U+2F46 “无”.

If it were a normal CJK character, according to the principle of unification, it should be unified; But, it is not a normal CJK character, it is a radical. The glyph “无” and “无” do have difference, whether the 乚 connects the 一 or not. This level of difference should not be considered as the same, that is because, even if U+2FA7 “長” and U+2FC1 “鬼” exist, it does not prevent the existence of U+2ED1 “長” and U+2EE4 “鬼”. The difference between the three pairs of glyphs are all the details of stroke cohesion, which is actually from the difference between 舊字形 and 新字形 in the mainland of China.

If I remember correctly, “鬼” was also considered as a duplication of “鬼” at that time, but later someone explained the difference and the necessity to encode them separately, which made them actually encoded separately without unification.

In this case, to fill the hole at U+2E9A with the glyph “无” can yet be regarded as a good choice. The UnicodeData of this character could be:

2E9A;CJK RADICAL NOT;So;0;ON;;;;;N;;;;;

Filling the hole could make the block neater, and more similar as the other CJK blocks. What is more, this operation can reduce the misunderstanding that “In the early years, the already defined characters could be deleted” , which is inconsistent with Core Spec.

For conclusion, I strongly propose to add the character back. But if everything else fails, at any rate, please add a comment in the CodeCharts as the red part below (like the comments in the Mathematical Alphanumeric Symbols block):

2E99	攴	CJK RADICAL RAP ·form used on right side → 6535 攴
2E9A	■	<reserved> → 2F46 无
2E9B	无	CJK RADICAL CHOKE → 65E1 无

(End of document)