

Response to IRG N2716 "Request to Remove Huìmìzì from Unicode Draft"

Andrew West, Ken Lunde, Eiso Chan, Huáng Jùnliàng

7th September 2024

The Daoist-usage characters proposed by the UK for IRG Working Set 2021 have been discussed by IRG experts multiple times over the four years of IRG review of WS2021, and the consensus of IRG experts is that they are suitable for encoding as CJK unified ideographs, and should be included in the repertoire of CJK Unified Ideographs Extension J. Only a single IRG individual expert, Mr. Wang Xieyang 王谢杨, was persistently and loudly opposed to their encoding, giving a variety of inconsistent and changing reasons as to why they should not be encoded.¹ However, his opinions were not supported by a majority of IRG experts, and no IRG member bodies formally asked for the removal of Daoist-usage characters proposed by the UK. There is absolutely no justification for overruling the recommendations of IRG ([M62.05](#)) and WG2 ([M71.14](#)) on the basis of the subjective opinions of an individual who has not been involved in the IRG review process.

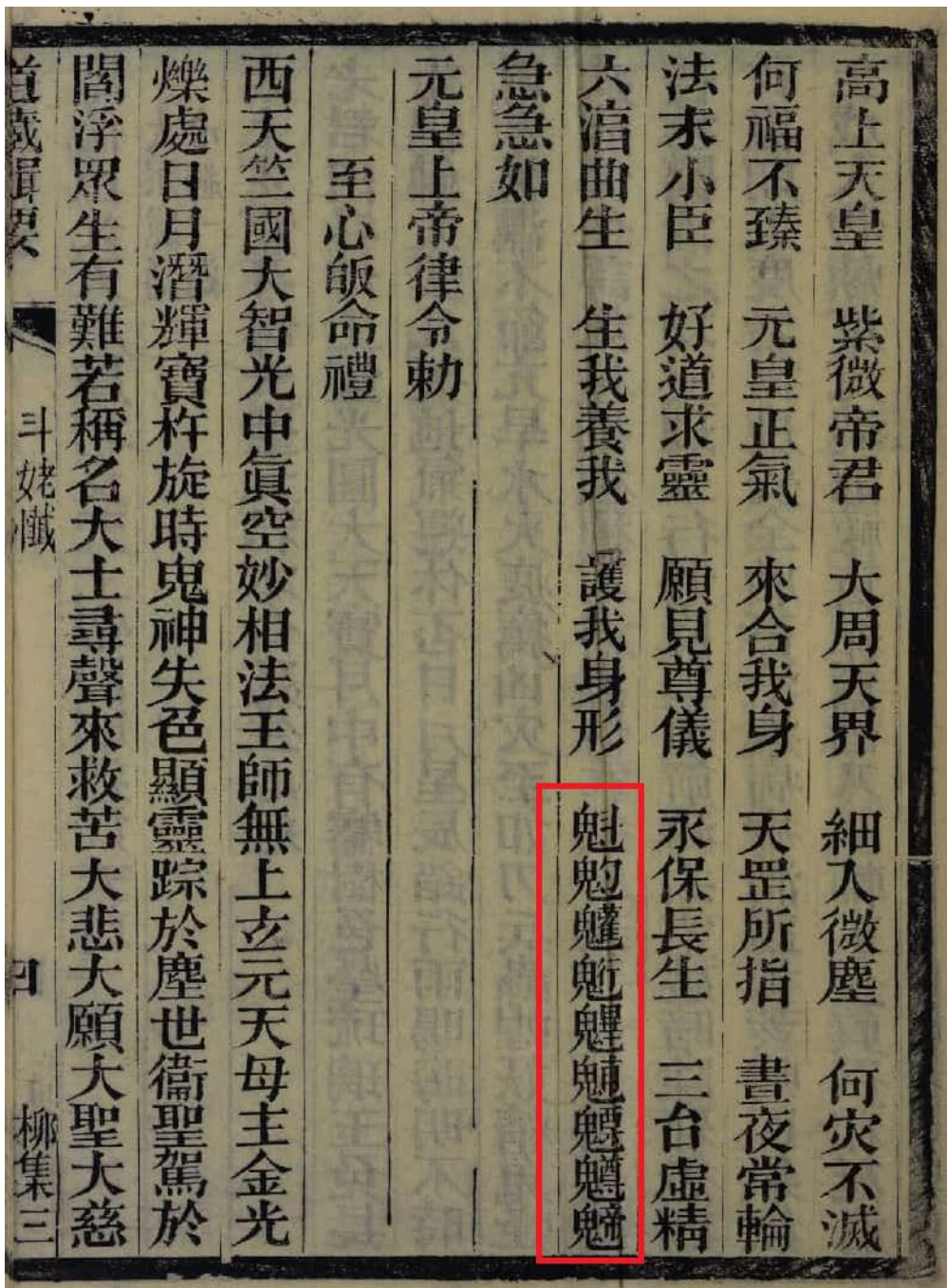
It should be noted that while there are many unencoded Daoist-usage characters, the Unicode and ISO/IEC 10646 standards already include a significant number of Daoist-usage characters within the URO and Extensions A and B, e.g. U+9B41 魁, U+9B52 魑, U+9B53 魍, U+4C22 魃, U+29C83 魃, U+29CD0 魃, U+29D44 魃 which are used as the Daoist names of the seven stars of the Big Dipper (北斗七星). The inclusion of these characters has never been a cause of concern for implementers of the standards or for font developers, and there is absolutely no reason to suppose that encoding additional Daoist-usage characters would be in any way destabilizing to the Unicode and ISO/IEC 10646 standards, or more burdensome for font developers than any other set of CJK ideographs.

Our brief responses to the four rationales against encoding given on page 2 of IRG N2716 are provided below.

¹ We note that Wang Xieyang recently submitted a proposal to the UTC ([L2/24-179](#)) to encode a set of 23 Daoist characters attested in modern Daoist books (published 1913 through 2004), so we assume that he no longer opposes the encoding of Daoist characters submitted by the UK. At the very least it would be unreasonable and hypocritical for him to oppose the Daoist characters in CJK Ext. J, which have a longer history of usage and have more stable glyph forms than the 23 characters in his proposal (because the sources for the 23 characters only emerged during the late 19th or early 20th century, we are not sure whether some of these characters may be variants of characters which appear in older sources).

• Huìmìzì are symbols, rather than normal hanzis.

This is simply not true. The vast majority of Daoist-usage characters are constructed in the normal manner from common CJK components, and conform to user expectations of *hanzi*. In UK-provided evidence showing their usage in running text, the Daoist-usage characters are indistinguishable in form and function from encoded CJK ideographs occurring in the same context. For example, in the text shown below ([懺法大觀](#)), the unencoded characters 魁 ([UK-20787](#)) and 魘 ([UK-30067](#)) are appended to the Daoist names for the seven stars of the Big Dipper 魁魘魘魘魘魘魘, and there is clearly no difference between the seven encoded and the two unencoded characters.



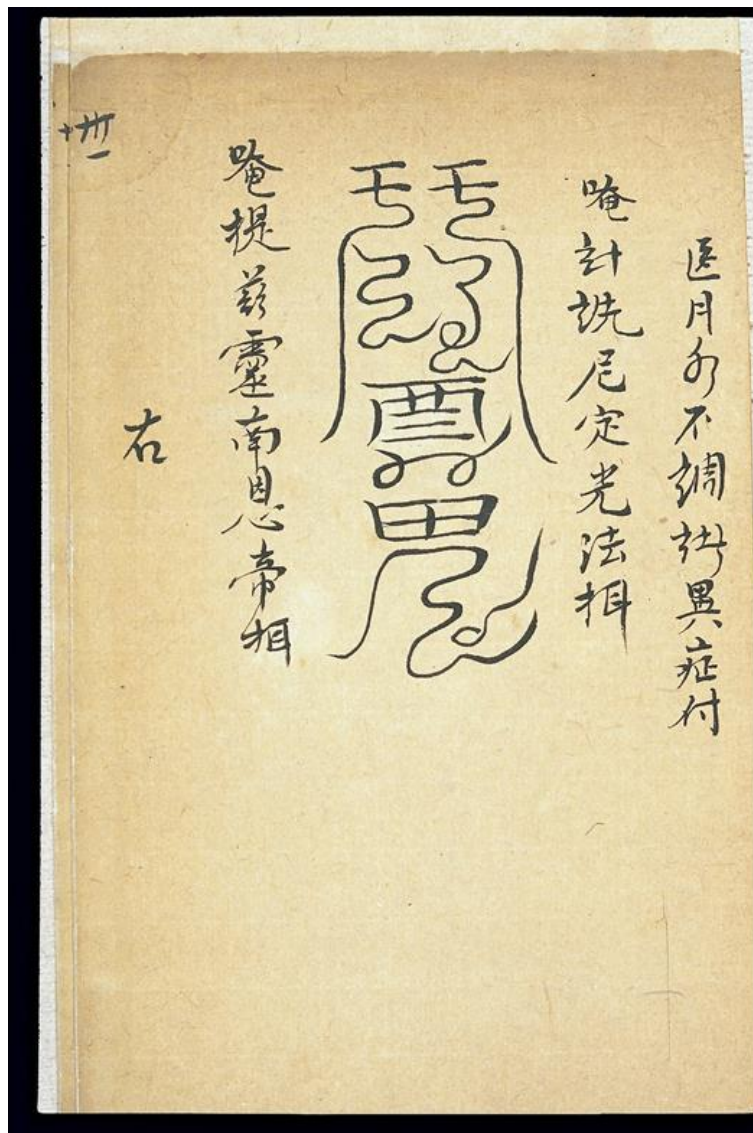
In other UK-provided evidence that shows Daoist-usage characters separately within sets of related characters, the proposed characters occur together with other already-encoded Daoist-usage CJK ideographs, and the proposed and encoded characters are indistinguishable in form and function. This is illustrated in the example below, where the encoded characters 魁 (in the centre of the diagram) and 魁魁魁魁魁魁魁 are the Daoist names for the seven stars of the Big Dipper, and the proposed characters 魁 (UK-20795) and 魁 (UK-20796) are the names of its two adjutant stars.



- Huìmìzì are only small fraction of all fúlù symbols and encoding them does not mean full digitalization of all fúlù.

The Chinese word *fúlù* 符箓 normally refers to Daoist talisman, such as the example shown below ([Wikimedia Commons](#) © ⓘ CC BY 4.0). It is disingenuous to conflate Daoist-usage ideographic characters with Daoist talismanic drawings which no-one considers to be suitable for encoding as characters. Although some of the proposed Daoist characters may be used in conjunction with a Daoist talisman for the incantation, they may also be used by themselves, unconnected to a talisman.

The existence of many thousands of talismanic *fúlù* should have absolutely no bearing on whether Daoist-usage ideographic characters should be encoded. Indeed, the fact that the Daoist-usage ideographic characters proposed by the UK in WS2021 are only a fraction of all unencoded Daoist-usage ideographic characters should not be a barrier to encoding. There being many other not-yet-encoded characters is no reason not to encode a subset, or else we would have stopped encoding any new CJK unified ideographs many years ago.



• **Throughout the history huìmìzì have almost always been created in large amounts by individuals casually. New huìmìzì would keep appearing ad infinitum, and encoding large quantities of rare huìmìzì is a waste of public resources. It is much better to use PUA characters or images instead.**

Of course individuals could casually create any number of new characters, but that does not mean that any of these novel characters would be candidates for encoding as CJK unified ideographs. IRG has a Principles and Procedures document ([IRG N2652](#)) which prohibits the encoding of novel characters created by random individuals, so this really should not be a concern.

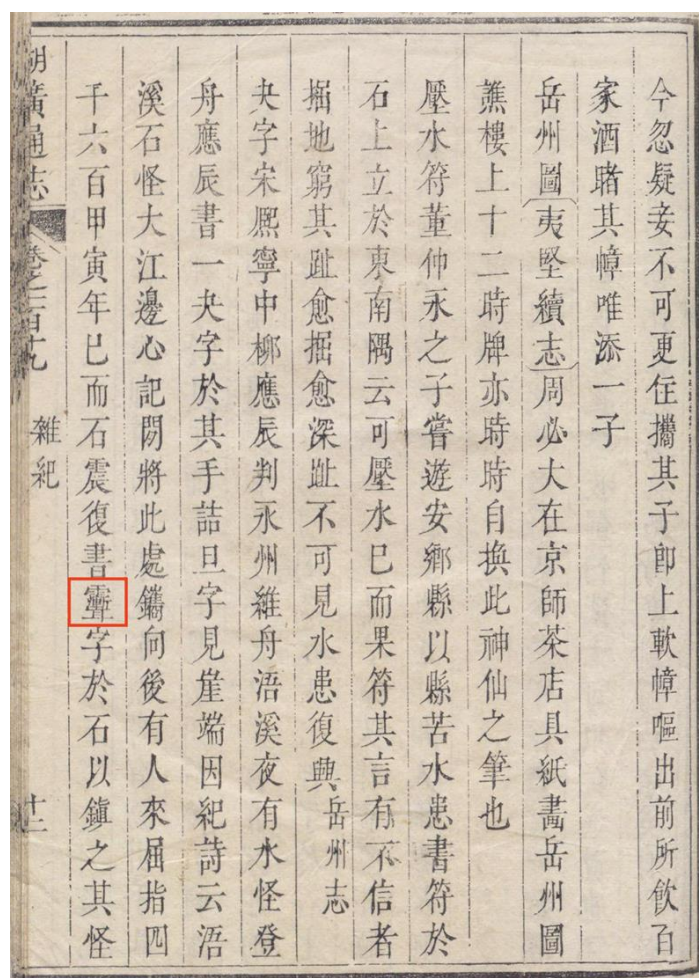
From a UK perspective, we only propose Daoist-usage characters that are attested in pre-modern printed sources, in most cases where the proposed characters are required for digitization of these texts. We need not go into the well-known details of why PUA characters or images are totally inappropriate for text digitization projects.

For IRG Working Set 2024, the UK and TCA have both submitted a large number of Daoist-usage ideographic characters. These are primarily sourced to the Zhengtong edition of the Daoist Canon printed between 1445 and 1447 (明正統道藏), which is an extremely important and authoritative source for Daoist texts. Characters included in this source are as equally required for encoding as the thousands of Buddhist-usage characters attested in the Buddhist canon. The Daoist-usage characters submitted by the UK for WS2024 are by no means a complete set of required characters used in the Zhengtong Daoist Canon, but are a carefully selected subset of those characters for which (in the vast majority of cases) multiple attestations have been found, both in the original woodblock edition of the Zhengtong Daoist Canon and in the modern typeset edition of the Daoist Canon (*Zhōnghuá Dàozàng* 中華道藏, 北京: 華夏出版社, 2004).

We have absolutely no expectation or desire that characters only attested in manuscript sources, modern-invented characters, or characters only listed in code charts showing PUA repertoires of specialist fonts should be encoded in the Unicode and ISO/IEC 10646 standards.

- The source *Guǎngchéng Yìzhì: Tiěguàn Shīshí Jí* 廣成儀制·鐵鎖施食集（清宣統二年刊本）UK used for huìmìzì in WS2021, is unreliable.

The UK submission for WS2021 used two separate primary sources for Daoist-usage characters, many of which occur in both sources. Whether or not these sources are reliable is a matter of opinion, but the characters attested in these two sources have been individually reviewed and accepted by IRG. During the review process we have weeded out some characters, and corrected the glyph forms of some other characters. Moreover, additional evidence from other sources has been provided in many cases. For example, for [UK-20679](#) 𪛗, which N2716 claims is a “suspicious glyph and usage”, six additional pieces of evidence showing usage of this character have been provided from various woodblock printed sources. For [UK-20785](#) 𪛗 (also submitted by China as GKJ-00998), [UK-20786](#) 𪛗, [UK-20787](#) 𪛗, and [UK-20788](#) 𪛗 (also submitted by China as GKJ-01004), which occur together in the incantation 𪛗𪛗𪛗𪛗𪛗𪛗 (𪛗 and 𪛗 are already encoded), there are altogether five separate pieces of evidence from five different woodblock printed texts, and three additional attestations for 𪛗 in other contexts have also been provided. N2716 claims that [UK-20701](#) 𪛗 (the Daoist name for the Purple Star 紫微星) is a “very unstable glyph”, but additional evidence has been provided in the ORT that shows this character with the same glyph form used in a 1733 edition of the *Huguang Provincial Gazetteer* 湖廣通志, so it is obviously not an unstable glyph.



Even in cases where no additional evidence has been found, the proposed characters are obviously suitable for encoding on the basis of the evidence provided. For example, N2716 claims that [UK-20698](#) 魑 and [UK-20699](#) 魍 are “unclear” and “rarely used” as rationales for not encoding. While it is true that these two ideographs are rarely used, that is also the case for the vast majority of characters in CJK Ext. J, so rarity of usage should not be a reason not to encode. As to the supposed unclearness of these two characters, the reader can judge for themselves from the evidence image shown below. Note how 魑 and 魍 are used in the same context as the encoded characters 魑 and 魍 and other common CJK ideographs.

