

ISO/IEC JTC1/SC2/WG2 N3196R2

Title	Proposal to Disunify U+4039
Source	Andrew West and John Jenkins
Document Type	Expert Contribution
Date	2007-05-01

1. Summary

The character U+4039 is a unification of two different glyphs, one written as 𦉳 (i.e. mù 目 radical and jiā 夾 phonetic) and one written as 𦉴 (i.e. mù 目 radical and shǎn 𦉴 phonetic). It is the former of these two glyph forms that is used to represent this character in the Unicode code charts and in most CJK fonts.

The glyph form of U+4039 in various sources are shown below :

Row/Cell	C	J	K	V
Hex Code	G-Hanzi-T	Kanji	Hanja	ChuNom
064/057	𦉳	𦉴		
4039	3-5952 3-5750	4-3946 4-2538		

ISO/IEC 10646:2003 page 307

04039 34590 34590 0809.030 42490.020 目 7	G3-5952 T4-3946 KP1-5E34	𦉳 𦉴
---	------------------------------------	--------

Super CJK Version 14.0 page 1049

𦉳
4039

The Unicode Standard 5.0 code charts page 301

1	1	1	0	0	0	1	81	璠	璣	璈
1	1	1	0	0	1	0	82	睨	𦉳	𦉴
1	1	1	0	0	1	1	83	穡	穠	穌

JIS X 0213:2000 Plane 2 Row 82 Col.2 (J4-7222)

There is also a simplified form of the *jiā* phonetic glyph (U+25174 𦉳), as well as two compatibility ideographs that are canonically equivalent to U+4039 : U+FAD4 𦉳 and U+2F949 𦉳. The situation is summarised in the table below :

Code Point	Character	Source References (from ISO/IEC 10646:2003 Amd.1)
4039	𦉳	G3-5952 T4-3946 J4-7222 H-98E6 KP1-5E34
FAD4	𦉳	KP1-5E2B
25174	𦉳	G_HZ
2F949	𦉳	T6-4B7A

We believe that the two glyph forms of U+4039 (𦉳 and 𦉳) are non-cognate, and so, according to the rules for CJK unification (see ISO/IEC 10646:2003 Annex S, S.1.1), should not have been unified. In accordance with Annex I ("Guideline for handling of CJK ideograph unification and/or disunification error") of the **Principles and Procedures** (N3102) we are requesting that a new character be encoded for the 𦉳 (*shǎn*) form of U+4039.

Note The misunification of U+4039 was first noted in IRG N855 (see also IRG N858), where it is included in a list of 305 unencoded characters in *Hanyu Da Zidian* 漢語大字典 as one of only four characters that were "unified by mistake" :

- U+385B 幌 vs. 巾 兪 兪 [Kangxi definition : 同幌 "same as U+3843"]
- U+393A 慌 vs. 忪 兪 [Kangxi definition : 慌字之譌 "mistake for U+393A"]
- U+4039 𦉳 vs. 𦉳
- U+5B90 𦉳 vs. 𦉳 互

However, only in the case of 𦉳/𦉳 is there a strong case to be made for disunification and the encoding of a new character.

2. Evidence for Disunification

The character written with the *jiā* 夾 phonetic is used as an alternate form for a number of different characters, and has the following pronunciations and meanings :

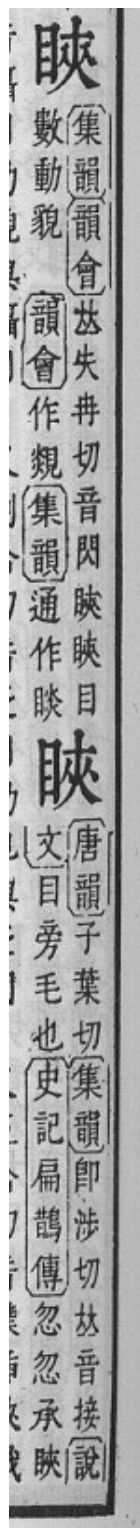
- *jiá* = 𦉳 *jiá* "to close one's eye(s)"
- *jié* = 𦉳 *jié* "eyelash"
- *shè* = 𦉳 *shè* "to blink"
- *zhǎ* = 𦉳 *zhǎ* "to blink"
- *yà* in the compound 𦉳 𦉳 *shà yà* "teasingly"

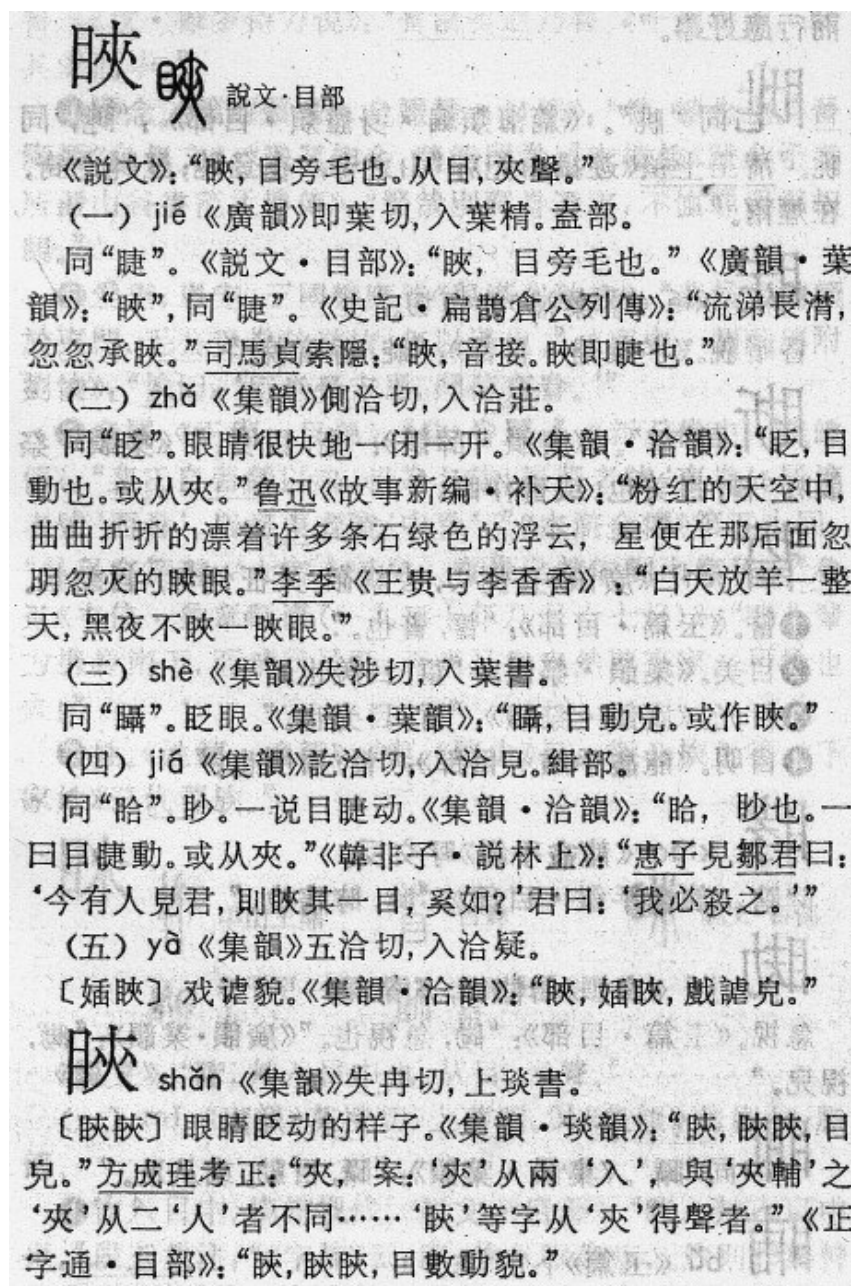
On the other hand the character written with the *shǎn* 𦉳 phonetic has only a single pronunciation and meaning :

- *shǎn* = 𦉳 *shǎn* "to blink" or "to twinkle" (of the eye)

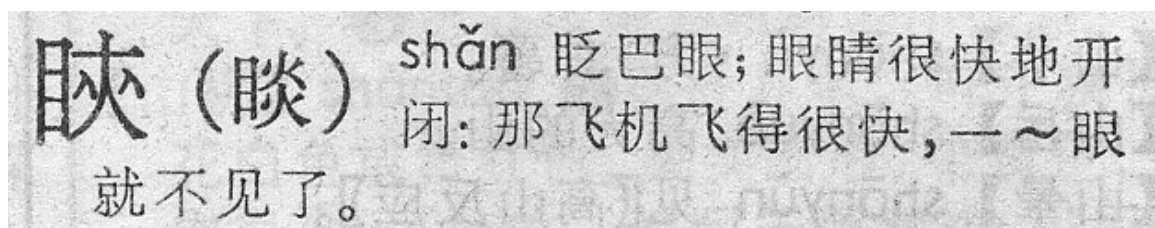
Whilst there is some semantic overlap between the two characters, it is clear that the characters are not cognate, and that the graphic similarity between them is coincidental. In all the dictionaries that we have consulted these two characters are lexically distinct, and neither is ever used as a variant form of the other. Note that the quotation by the early 19th century scholar Fang Chengui 方成珪 given in the entry for 眅 (*shǎn*) in *Hanyu Da Zidian* 漢語大字典 (Great Dictionary of Chinese Characters) explicitly states that the phonetic for this character is *shǎn* 爽, not *jiā* 夾.

The lexical separation of the characters 眅 (*jiá, jié, shè, zhǎ, yà*) and 眅 (*shǎn*) is clearly demonstrated by the examples below, taken from various authoritative modern and traditional dictionaries.

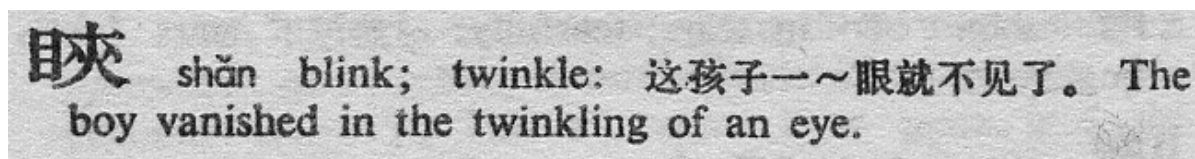




Hanyu Da Zidian 漢語大字典 (Sichuan Cishu Chubanshe, 1986-1990) vol.4 p.2490.



Xiandai Hanyu Cidian 現代漢語詞典 (Shangwu Yinshuguan, 2nd ed., 1983) p.998.



Han-Ying Cidian 漢英詞典 (Shangwu Yinshuguan, 1980) p.594.

眇¹ [jié<廣韻> 卽葉切,入葉,精。] 眼睫毛。《史記·扁鵲倉公列傳》:“言未卒,因噓唏服臆,魂精泄橫,流涕長潛,忽忽承眇。”司馬貞索隱:“音接。眇卽睫也。承眇,言淚恒垂以承於睫也。”

眇² [jiǎo<集韻> 訖洽切,入洽,見。] 眼睛閉上。《韓非子·說林上》:“惠子見鄒君曰:‘今有人見君,則眇其一目,奚如?’君曰:‘我必殺之。’”王煥鑣注:“眇,閉目;眇其一目,故意閉一目,是輕蔑人的態度。”清錢謙益《復吳江潘力田書》:“杜詩非易注之書,注杜非聊爾之事,固不妨慎之又慎,精之又精,終不應草次裨販,冀幸舉世兩目盡眇,而以爲予雄也。”

眇³ [zhǎo<集韻> 側洽切,入洽,莊。] 眨眼; 上下眼皮很快地開閉。魯迅《野草·秋夜》:“他仿佛要離開人間而去,使人們仰面不再看見。然而現在却非常之藍,閃閃地眇着幾十個星星的眼,冷眼。”艾蕪《端陽節》九:“張大爺光着頭赤着足尾着跑來。老三不禁笑了眇一眇眼睛。”

眇 [shǎn<集韻> 失冉切,上琰,書。] 眼很快地開閉。茅盾《賽會》二:“說不到三句話,他就搔着頭皮,眇着一隻眼。”趙樹理《李有才板話》一:“這人的相貌不大好看,臉像個葫蘆瓢子,說一句話眇十來次眼皮。”徐遲《地質之光》:“說到這里,白髮蒼蒼的李四光眇眇眼睛,笑了一笑,輕輕撥動他桌上一個地球儀,一下子使小小寰球急速地旋轉了起來。”

Hanyu Da Cidian 漢語大詞典 (Hanyu Dacidian Chubanshe, 1986-1993) vol.7 pp.1221-1222.

眇¹. jié 卽葉切,入,葉韻,精。
𠂔

⊖目旁毛。同“睫”。史記一〇五扁鵲傳:“流涕長潛,忽忽承眇。”⊖閉目。韓非子說林上:“今有人見君,則眇其一目,奚如?”

2. zhǎo 集韻 側洽切,入,洽韻。
𠂔

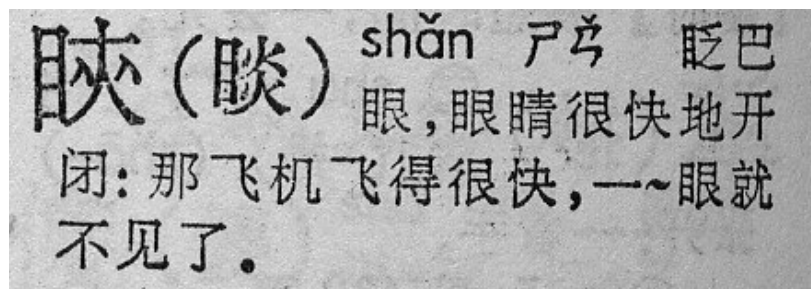
⊖目動。同“眨”。見集韻。

眇 shǎn 集韻 失冉切,上,琰韻。
𠂔

見下。

【眇眇】 眨目貌,指目頻頻動。見集韻。

Ci Yuan 辭源 (Shangwu Yinshuguan, compact ed., 1988) p.1202.



Xinhua Zidian 新華字典 (Shangwu Yinshuguan, 5th ed., 1979) p.398.

3. Need to Encode a New Character

Although the character 眅 (*shǎn*) is not in every day use, it is far from obscure, and is more commonly encountered than the character 眅 (*jiá*). Not only does 眅 (*shǎn*) occur in almost all modern Chinese dictionaries, but it is present in smaller dictionaries such as *Han-Ying Cidian* 漢英詞典 (Chinese-English Dictionary), *Xinhua Zidian* 新華字典 (New China Dictionary) and *Xiandai Hanyu Cidian* 現代漢語詞典 (Modern Chinese Dictionary) which do not have an entry for 眅 (*jiá*). The relative importance of the character 眅 (*shǎn*) can be judged from the fact that it is included in *Han-Ying Cidian* 漢英詞典 (Chinese-English Dictionary), which only has entries for somewhat over 6,000 Chinese characters.

The character is mostly used in modern literature as an alternative character for 眅 (*shǎn* ("to blink" or "to twinkle"). For example, the entry for this character in *Hanyu Da Cidian* 漢語大詞典 (Great Dictionary of Chinese Words) shown above quotes examples of its usage by famous modern authors such as Mao Dun 茅盾 (1896-1981), Zhao Shuli 趙樹理 (1906-1970) and Xu Chi 徐遲 (1914-1996). The character can also be found in pre-modern literature, such as the example below from the famous collection of fantastic tales by Pu Songling 蒲松齡 (1640-1715), *Liaozhai Zhiyi* 聊齋志異 (Strange Stories from a Chinese Studio):

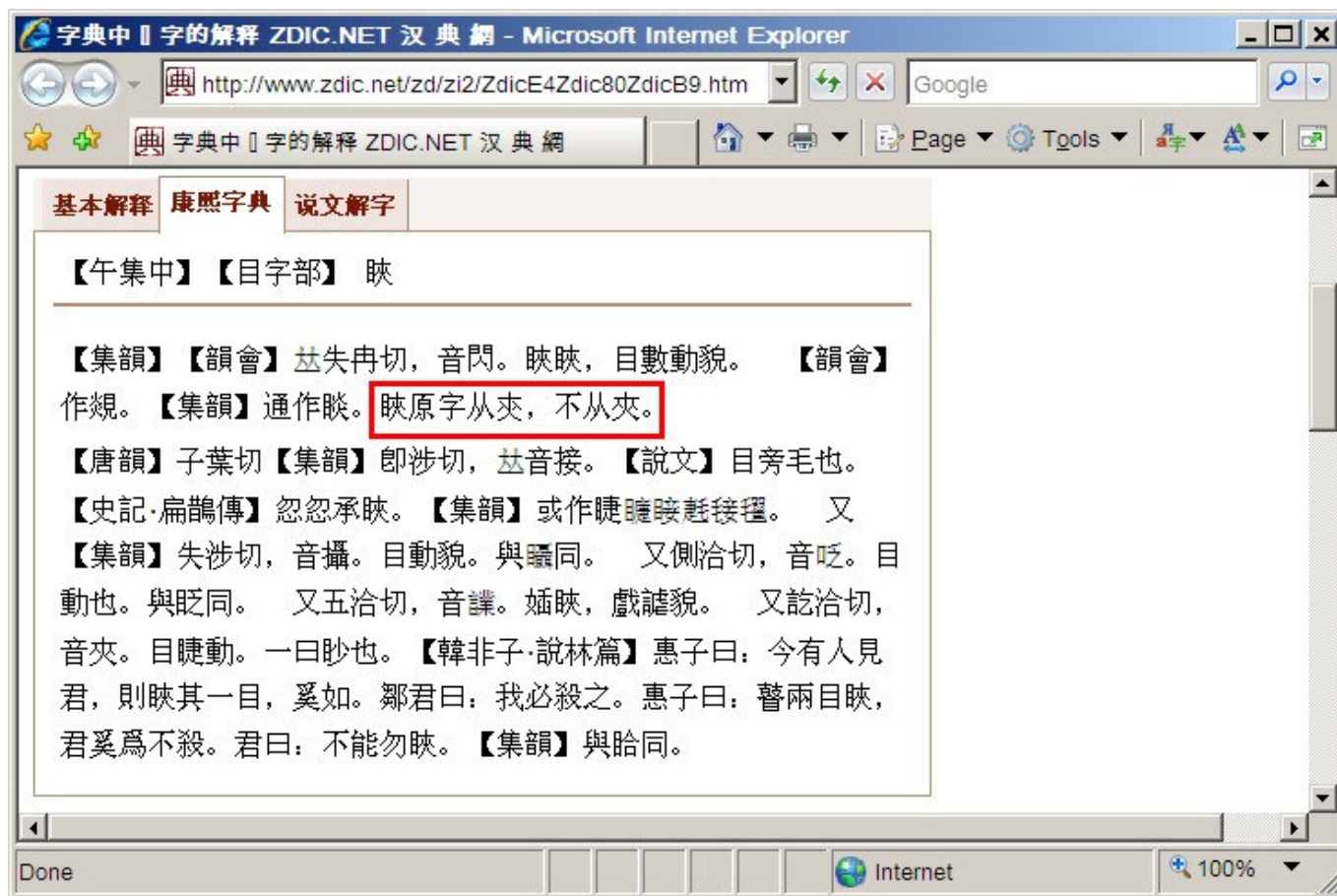
石乃即刻治任如汴，詢至玄帝觀，則赤城歸未久。入而參之，便問：“何來？”石視厨下一老狐，孔前股而系之，笑曰：“弟子之來，為此老魅。”赤誠詰之，曰：“是吾岳也。”因以實告。道士謂其狡詐，不肯輕釋。固請，乃許之。石因備述其詐，狐聞之，塞身入炷，似有慚狀。道士笑曰：“彼羞惡之心，未盡亡也。”石起，牽之而出，以刀斷索抽之。狐痛極，齒齟齬然。石不遽抽，而頓挫之，笑問曰：“翁痛之，勿抽可耶？”狐睛眅閃，似有慍色。既釋，搖尾出觀而去。

狐睛眅閃

Liaozhai Zhiyi Pingshang Dacheng 聊齋志異評賞大成 (Lijiang Chubanshe 濰江出版社, 1992) p.2120.

Clearly there is a need to be able to represent the character 眅 (*shǎn*) in electronic texts, and this need is not currently met. The fact that a compatibility ideograph with the same glyph shape (U+FAD4) is encoded does not help most users, as their data may undergo normalization during interchange, in which case the compatibility ideograph U+FAD4 will be normalized to U+4039.

The problems that users currently have in representing the character 眅 (*shǎn*) maybe illustrated by this screen shot of a web version of the Kangxi Dictionary :



<<http://www.zdic.net/zd/zi2/ZdicE4Zdic80ZdicB9.htm>>

Although this online version of the Kangxi Dictionary incorrectly conflates the two entries for the characters 眅 (*shǎn*) and 眅 (*jiá*), and uses U+4039 for both characters, it recognises that the two characters are different, but is unable to represent this difference at the encoding level. Instead the editor has had to resort to interpolating an apologetic note into the text that states 眅原字从夾，不从夾。 "this character is originally written with the *shan* phonetic not the *jia* phonetic".

4. Proposed Solution

4.1 Disunification

The examples given in Section 2 (Evidence for Disunification) clearly demonstrate that the characters 眅 (*shǎn*) and 眅 (*jiá*) are a non-cognate pair. According to **Principles and Procedures (N3102) I.3, If [a] character pair is non-cognate (meanings are different), that pair of characters is TO BE DISUNIFIED.** Annex I further states that in the case of a *to be disunified* error :

The ideographs to be disunified should be disunified and should be given separate code positions as soon as possible (disunification in some sense, and character name change in some sense also). These ideographs will have two separate glyphs and two separate code positions. One of these ideographs will stay at its current encoded position. The other one will have a new glyph and a new code position.

As to which character should be associated with the existing code point and which should be associated with the new code point, Annex I mandates :

In the interest of synchronization between ISO/IEC 10646 and the Unicode standard, the ideograph with the glyph shape that is similar to the glyph that is published in the “Unicode Charts” will continue to be associated with its current code position. For the ideographs outside the BMP, the glyph shape in ISO/IEC 10646 and the Unicode Charts are identical and will be used with its current code position.

As the Unicode charts use the 𠄎 (*jiá*) glyph form, the following solution should be implemented :

1. The character U+4039 will retain the 𠄎 (*jiá*) glyph form
2. A new character will be encoded for the 𠄎 (*shǎn*) glyph form (U+9FC3 suggested)

4.2 Source References

U+4039 has the following source references :

- G3-5952
- T4-3946
- J4-7222
- H-98E6
- KP1-5E34

We believe that G3-5952, J4-7222, H-98E6 and KP1-5E34 refer to the 𠄎 (*jiā*) glyph form, whereas T4-3946 refers to the 𠄎 (*shǎn*) glyph form. The source reference for the compatibility ideograph U+2F949 (T6-4B7A) refers to the 𠄎 (*jiā*) glyph form, whereas the source reference for the compatibility ideograph U+FAD4 (KP1-5E2B) refers to the 𠄎 (*shǎn*) glyph form.

Therefore, the source references G3-5952, J4-7222, H-98E6, KP1-5E34 and T6-4B7A should be associated with the original character (U+4039), and the source references T4-3946 and KP1-5E2B should be associated with the new character (U+9FC3).

4.2 Compatibility Ideographs

The two existing compatibility ideographs U+FAD4 𠄎 and U+2F949 𠄎 will continue to be canonically equivalent to U+4039. In the case of U+FAD4 this means that it will be canonically equivalent to the wrong character, but this is unavoidable. However, as U+FAD4 should no longer be needed after the encoding of the new character it can effectively be deprecated in favour of the new character U+9FC3 (indeed it may be prudent to formally deprecate U+FAD4 if the new character is encoded).

4.3 Simplification Issues

Encoding a new character for 𠄎 (*shǎn*) will have no effect on the simplified character U+25174 𠄎, as there is no formal relationship between it and U+4039 (the UniHan database does not currently give a kTraditionalVariant mapping for U+25174). Note that although U+965D 𠄎 simplifies to U+9655 𠄎, this simplification is anomalous, and U+25174 would be expected to only be the simplified form of U+4039 𠄎 (*jiā*) and not of the proposed new character U+9FC3 𠄎 (*shǎn*).

4.4 UniHan Data

The UniHan data for U+4039 is currently (Unicode 5.0.0) as follows :

```

U+4039 kCangjie BUK00
U+4039 kCantonese gap6 sip3 zip3
U+4039 kCheungBauer 109/07;BUK00;gap6,sip3
U+4039 kCheungBauerIndex 443.11 444.01
U+4039 kCihaiT 951.404
U+4039 kDefinition (same as U+7728 眨) to wink; (same as U+776B 睫) eyelashes,
having one eye smller than the other, joke; witticism; pleasantry; jest; fun;
(Cant.) to peep at; to blink, wink
U+4039 kFennIndex 57.01
U+4039 kHKSCS 98E6
U+4039 kHanYu 42490.020 42490.030
U+4039 kIRGHanyuDaZidian 42490.020
U+4039 kIRGKangXi 0809.030
U+4039 kIRG_GSource 3-5952
U+4039 kIRG_HSource 98E6
U+4039 kIRG_JSource 4-7222
U+4039 kIRG_KPSource KP1-5E34
U+4039 kIRG_TSource 4-3946
U+4039 kJIS0213 2,82,02
U+4039 kKPS1 5E34
U+4039 kMandarin JIA2 SHE4
U+4039 kMatthews 790
U+4039 kPhonetic 550
U+4039 kRSAdobe_Japan1_6 C+18191+109.5.7
U+4039 kRSUnicode 109.7
U+4039 kSBGY 538.34
U+4039 kSemanticVariant U+776B<kMatthews
U+4039 kTotalStrokes 12

```

If the disunification of U+4039 is accepted we would expect the UniHan data for the two characters to be as follows :

```

U+4039 kCangjie BUK00
U+4039 kCantonese gap6 sip3 zip3
U+4039 kCheungBauer 109/07;BUK00;gap6,sip3
U+4039 kCheungBauerIndex 443.11
U+4039 kCihaiT 951.308
U+4039 kDefinition (same as U+7728 眨) to wink; (same as U+776B 睫) eyelashes,
having one eye smller than the other, joke; witticism; pleasantry; jest; fun;
(Cant.) to peep at; to blink, wink
U+4039 kFennIndex 57.01
U+4039 kHKSCS 98E6
U+4039 kHanYu 42490.020
U+4039 kIRGHanyuDaZidian 42490.020
U+4039 kIRGKangXi 0809.030
U+4039 kIRG_GSource 3-5952
U+4039 kIRG_HSource 98E6
U+4039 kIRG_JSource 4-7222
U+4039 kIRG_KPSource KP1-5E34
U+4039 kIRG_TSource 6-4B7A
U+4039 kJIS0213 2,82,02
U+4039 kKPS1 5E34
U+4039 kMandarin JIE2 ZHA3 JIA2 SHE4 YA4
U+4039 kMatthews 790
U+4039 kPhonetic 550
U+4039 kRSAdobe_Japan1_6 C+18191+109.5.7
U+4039 kRSUnicode 109.7
U+4039 kSBGY 538.34
U+4039 kSemanticVariant U+776B<kMatthews
U+4039 kSimplifiedVariant U+25174
U+4039 kTotalStrokes 12

```

U+9FC3 kCangjie BUKOO
U+9FC3 kCantonese sim2
U+9FC3 kCheungBauer 109/07;BUKOO;
U+9FC3 kCheungBauerIndex 444.01
U+9FC3 kCihaiT 951.404
U+9FC3 kDefinition (same as U+7752 睽) to blink, twinkle
U+9FC3 kHanYu 42490.030
U+9FC3 kIRGHanyuDaZidian 42490.030
U+9FC3 kIRGKangXi 0809.020
U+9FC3 kIRG_GSource KX
U+9FC3 kIRG_KPSource KP1-5E2B
U+9FC3 kIRG_TSource 4-3946
U+9FC3 kKPS1 5E2B
U+9FC3 kMandarin SHAN3
U+9FC3 kPhonetic 1197
U+9FC3 kRSUnicode 109.7
U+9FC3 kTotalStrokes 12
