ISO/IEC JTC1/SC2/WG2 N3196R

Title	**Proposal to Disunify U+4039**
Source	Andrew West and John Jenkins
Document Type	Expert Contribution
Date	8th January 2007, revised 18th March 2007

Background

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

Super CJK Version 14.0 page 1049

There is also a simplified form of the *jiā* phonetic glyph (U+25174 睞), as well as two compatability ideographs that are canonically equivalent to U+4039 : U+FAD4 睞 and U+2F949 睞.

We believe that the two characters 睞 (*jiā* phonetic) and 睞 (*shǎn* phonetic) are lexically distinct, and so should not have been unified according to the rules for CJK unification (see ISO/IEC 10646:2003 Annex S, S.1.1).

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

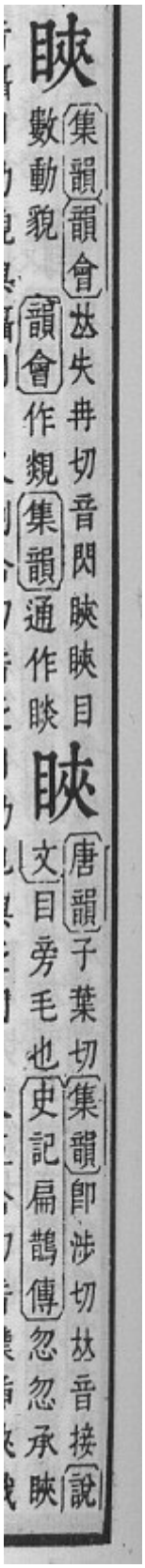
- *jié* = 睫 *jié* "eyelash"
- *zhǎ* = 眨 *zhǎ* "to blink"
- *jiá* "to close one's eye(s)"
- *shè* = 䁖 *shè* "to blink"
- *yà* in the compound 婚睞 *shà yà* "teasingly"

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

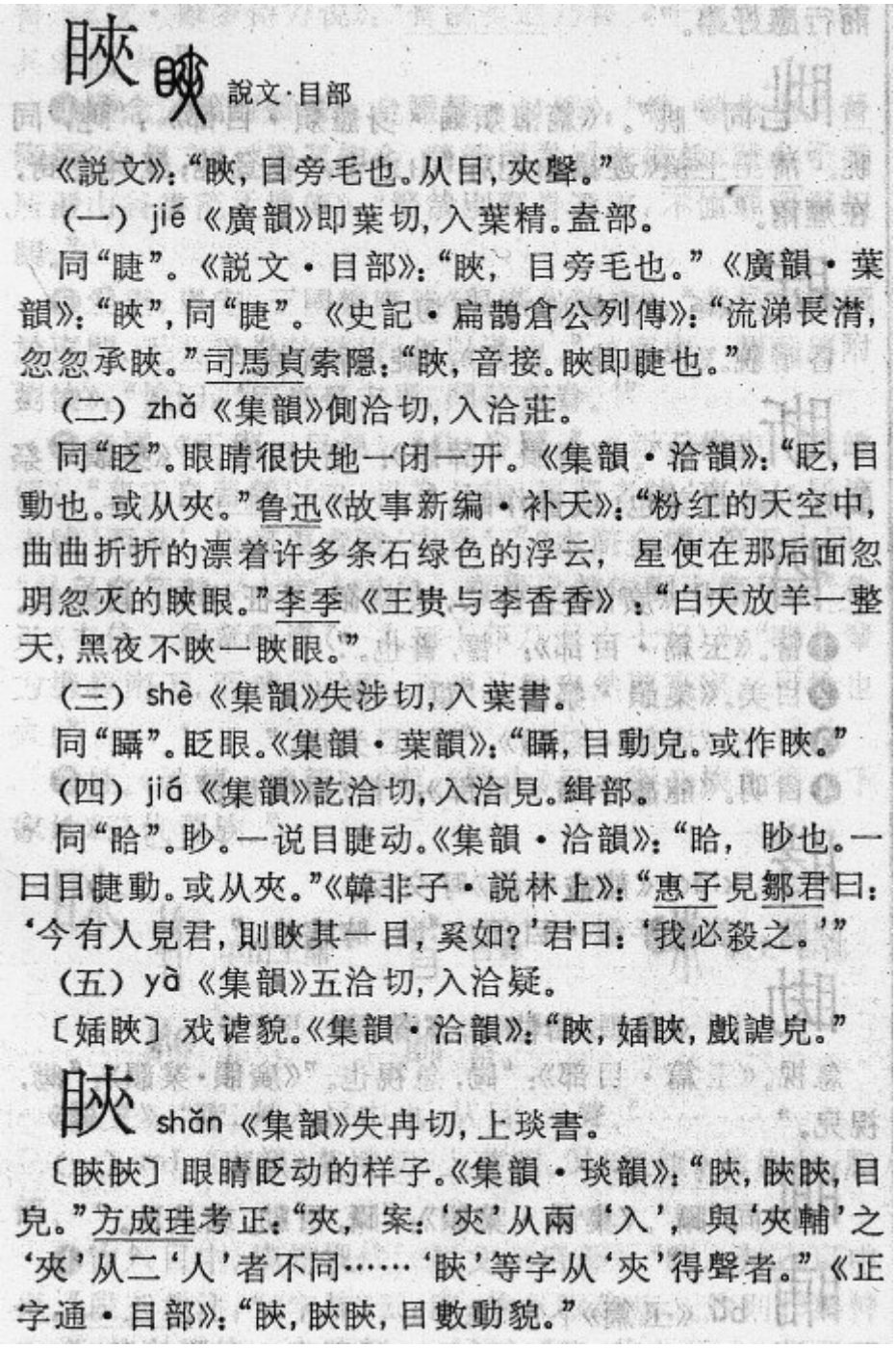
- *shǎn* = 睒 *shǎn* "to blink" or "to twinkle"

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. The lexical separation of these two characters is clearly demonstrated by the examples below, taken from various authoritative modern and traditional dictionaries. In smaller modern Chinese dictionaries such as *Xinhua Zidian* 新華字典 and *Xiandai Hanyu Cidian* 現代漢語詞典, the character with the *jiā* 夾 phonetic (i.e. the representative

glyph form of U+4039) does not occur, and only the character with the *shǎn* 夾 phonetic is found.



Kangxi Zidian 康熙字典 (Zhonghua Shuju, 1958) p.809.



睞 說文·目部

《說文》:"睞,目旁毛也。从目,夾聲。"

(一) jié 《廣韻》即葉切,入葉精。盍部。

同"睫"。《說文·目部》:"睞,目旁毛也。"《廣韻·葉韻》:"睞",同"睫"。《史記·扁鵲倉公列傳》:"流涕長潸,忽忽承睞。"司馬貞索隱:"睞,音接。睞即睫也。"

(二) zhǎ 《集韻》側洽切,入洽莊。

同"眨"。眼睛很快地一闭一开。《集韻·洽韻》:"眨,目動也。或从夾。"鲁迅《故事新编·补天》:"粉红的天空中,曲曲折折的漂着许多条石绿色的浮云,星便在那后面忽明忽灭的睞眼。"李季《王贵与李香香》:"白天放羊一整天,黑夜不睞一睞眼。"

(三) shè 《集韻》失涉切,入葉書。

同"瞬"。眨眼。《集韻·葉韻》:"瞬,目動皃。或作睞。"

(四) jiá 《集韻》訖洽切,入洽見。緝部。

同"眙"。眇。一说目睫动。《集韻·洽韻》:"眙,眇也。一曰目睫動。或从夾。"《韓非子·說林上》:"惠子見鄒君曰:'今有人見君,則睞其一目,奚如?'君曰:'我必殺之。'"

(五) yà 《集韻》五洽切,入洽疑。

〔媌睞〕戏謔貌。《集韻·洽韻》:"睞,媌睞,戲謔皃。"

睞 shǎn 《集韻》失冉切,上琰書。

〔睞睞〕眼睛眨动的样子。《集韻·琰韻》:"睞,睞睞,目皃。"方成珪考正:"夾,案:'夾'从兩'入',與'夾輔'之'夾'从二'人'者不同……'睞'等字从'夾'得聲者。"《正字通·目部》:"睞,睞睞,目數動貌。"

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

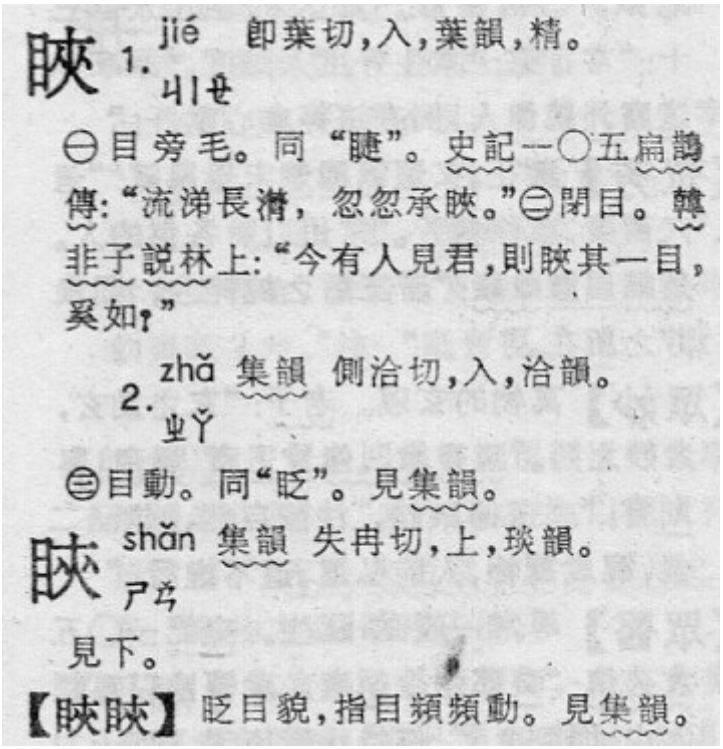
睫[1] [jié《广韵》即葉切，入葉，精。]眼睫毛。《史记·扁鹊仓公列传》："言未卒，因噓唏服臆，魂精泄横，流涕長潸，忽忽承睫。"司马贞索隐："音接。睫即睫也。承睫，言淚恒垂以承於睫也。"

睫[2] [jiá《集韵》訖洽切，入洽，見。]眼睛闭上。《韩非子·说林上》："惠子見鄒君曰：'今有人見君，則睫其一目，奚如？'君曰：'我必殺之。'"王焕镳注："睫，閉目；睫其一目，故意閉一目，是輕蔑人的態度。"清钱谦益《复吴江潘力田书》："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，冀幸舉世兩目盡睫，而以爲予雄也。"

睫[3] [zhǎ《集韵》側洽切，入洽，莊。]眨眼；上下眼皮很快地开闭。鲁迅《野草·秋夜》："他仿佛要离开人间而去，使人们仰面不再看见。然而现在却非常之蓝，闪闪地睫着几十个星星的眼，冷眼。"艾芜《端阳节》九："张大爷光着头赤着足尾着跑来。老三不禁笑了睫一睫眼睛。"

睫 [shǎn《集韵》失冉切，上琰，書。]眼很快地开闭。茅盾《赛会》二："说不到三句话，他就搔着头皮，睫着一只眼。"赵树理《李有才板话》一："这人的相貌不大好看，脸像个葫芦瓢子，说一句话睫十来次眼皮。"徐迟《地质之光》："说到这里，白发苍苍的李四光睫睫眼睛，笑了一笑，轻轻拨动他桌上一个地球仪，一下子使小小寰球急速地旋转了起来。"

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



睞 1. jié 即葉切，入，葉韻，精。

㊀目旁毛。同“睫”。史記一〇五扁鵲傳：“流涕長潸，忽忽承睞。”㊁閉目。韓非子說林上：“今有人見君，則睞其一目，奚如？”

2. zhǎ 集韻 側洽切，入，洽韻。

㊀目動。同“眨”。見集韻。

睞 shǎn 集韻 失冉切，上，琰韻。

見下。

【睞睞】眨目貌，指目頻頻動。見集韻。

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

Proposed Solution

From the above examples it is clear that the two characters 睞 and 睞 are distinct and need to be encoded separately. There are two possibilities :

Solution A

The existing U+4039 is identified with the *jiā* phonetic character 睞

A new character U+XXXX is encoded for the *shǎn* phonetic character 睞

Solution B

The existing U+4039 is identified with the *shǎn* phonetic character 睞

A new character U+XXXX is encoded for the *jiā* phonetic character 睞

With regard to the simplified character U+25174 睞, it does not matter whether Solution A or Solution B is accepted 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 陝 simplifies to U+9655 陕, this simplification is anomalous, and U+25174 would be expected to only be the simplified form of 睞 (*jiā* phonetic) and not of 睞 (*shǎn* phonetic).

Unfortunately the issue is complicated by the existence of two compatibility ideographs, both canonically equivalent to U+4039, but one with the *jiā* 夾 phonetic (U+2F949 睞) and one with the *shǎn* 夾 phonetic (U+FAD4 睞). The former should be equivalent to the *jiā* phonetic character and the latter should be equivalent to the *shǎn* phonetic character, but as they are both immutably equivalent to U+4039, if Solution A is accepted U+FAD4 will be cannonically equivalent to the wrong character, whereas if Solution B is accepted U+2F949 will be cannonically equivalent to the wrong character -- whichever solution is chosen one of the compatibility characters will end up with an incorrect canonical equivalence.

However, U+FAD4 was only encoded because of the need to distinguish the two characters (*jiā* phonetic and *shǎn* phonetic), so accepting Solution A will make U+FAD4 redundant, and it can effectively be deprecated in favour of the new character U+XXXX (indeed it may be prudent to formally deprecate U+FAD4 if Solution A is accepted). Thus, leaving U+FAD4 with an incorrect canonical equivalence should not have a significant negative impact.

A further advantage of Solution A over Solution B is that the vast majority of existing CJK fonts use the *jiā* phonetic glyph for U+4039, so keeping U+4039 as the *jiā* phonetic character will minimise the need to change existing fonts.

Therefore we believe that the best solution would be to encode the character *shǎn* 睞 (*mù* 目 radical and the *shǎn* 夾 phonetic) as a separate character in its own right in a future amendment to ISO/IEC 10646.

UniHan Data

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

```
U+4039 kCangjie BUKOO
U+4039 kCantonese gap6 sip3 zip3
U+4039 kCheungBauer 109/07;BUKOO;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 Solution A proposed above is accepted we would expect the UniHan data for the two characters to be as follows (where U+4039 is the *jiā* phonetic character and U+XXXX is the *shǎn* phonetic character) :

```
U+4039 kCangjie BUKOO
U+4039 kCantonese gap6 sip3 zip3
U+4039 kCheungBauer 109/07;BUKOO;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 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+XXXX kCangjie BUKOO
U+XXXX kCantonese sim2
U+XXXX kCheungBauer 109/07;BUKOO;
U+XXXX kCheungBauerIndex 444.01
U+XXXX kCihaiT 951.404
U+XXXX kDefinition (same as U+7752 睒) to blink, twinkle
U+XXXX kHanYu 42490.030
U+XXXX kIRGHanyuDaZidian 42490.030
U+XXXX kIRGKangXi 0809.020
U+XXXX kMandarin SHAN3
U+XXXX kPhonetic 1197
U+XXXX kRSUnicode 109.7
U+XXXX kTotalStrokes 12
```

N.B. Some of these fields need to be checked.